101 Rec'd PCT/PTO 19 JUN 1998

FORM PTO-1390 (REV. 5/93) U.S. Department of Commerce Patent and Trademark Office		Attorney's Docket Number 117-260
TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371		U.S. Application No. (if known, see 37 C.F.R. 1.5) 09/091538 (To Be Assigned)
International Application No. PCT/GB96/03221	International Filing Date 23 December 1996	Priority Date Claimed 21 December 1995

Title of Invention

NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES IN PATHOGENIC MYCOBACTERIA AND THEIR USE AS DIAGNOSTICS, VACCINES AND TARGETS FOR CHEMOTHERAPY

Applicant(s) For DO/EO/US

HERMON-TAYLOR et al

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information.

1. ☒ This is a **FIRST** submission of items concerning a filing under 35 U.S.C. 371.
2. ☐ This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.
3. ☒ This is an express request to begin national examination procedures (35 U.S.C. 371(f) at any time rather than delay examination until the expiration of the applicable time limit set in 35 U.S.C. 371(b) Articles 22 and 39(1).
4. ☒ A proper Demand for International Preliminary Examination was made by the 19th month from the earliest claimed priority date.
5. A copy of the International Application as filed (35 U.S.C. 371(c)(2)).
   a. ☐ is transmitted herewith (required only if not transmitted by the International Bureau).
   b. ☒ has been transmitted by the International Bureau.
   c. ☐ is not required, as the application was filed in the United States Receiving Office (RO/US).
6. ☐ A translation of the International Application into English (35 U.S.C. 371(c)(2)).
7. Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3)).
   a. ☐ are transmitted herewith (required only if not transmitted by the International Bureau).
   b. ☐ have been transmitted by the International Bureau
   c. ☐ have not been made; however, the time limit for making such amendments has **NOT** expired.
   d. ☐ have not been made and will not be made.
8. ☐ A translation of the amendments to the claims under PCT Article 19 (U.S.C. 371(c)(3)).
9. ☐ An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).
10. ☐ A translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).
11. The above checked items are being transmitted:
    a. ☐ before the 18th month publication.
    b. ☐ after publication and the Article 20 communication but before 20 months from the priority date.
    c. ☐ after 20 months.
    d. ☒ by 30 months and a proper demand for International Preliminary Examination was made by the 19th month from the earliest claimed priority date.
    e. ☐ after 30 months.

    Note: Petition to revive (37 CFR 1.137(a) or (b)) is necessary if 35 U.S.C. 371 requirements submitted (1) after 20 months and no proper demand for International Preliminary Examination was made by 19 months from the earliest claimed priority date, or (2) after 30 months and a proper demand for International Preliminary Examination was made by 19 months from the earliest claimed priority date.
12. At the time of transmittal, Amendments to the claims under Article 34
    a. ☐ are transmitted herewith (required only if not transmitted by the International Bureau).
    b. ☒ have been transmitted by the International Bureau
    c. ☐ have not been made; however, the time limit for making such amendments has **NOT** expired.
    d. ☐ have not been made and will not be made.
13. ☐ Certain requirements under 35 U.S.C. 371 were previously submitted by the applicant on _____ , namely:

14. ☒ An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
15. ☐ An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.
16. ☒ A **FIRST** preliminary amendment.
    ☐ A **SECOND** OR **SUBSEQUENT** preliminary amendment.
17. ☐ A substitute specification.
18. ☐ A change of power of attorney and/or address letter.

Attorney's Docket Number **117-260**

19. ☒ Other items or information: **International Search Report, Sequence Listing (Paper Form)**							
20. ☒ The following fees are submitted:						CALCULATIONS	PTO USE ONLY
BASIC NATIONAL FEE (37 CFR 1.492(a)(1)-(5) -- Search Report has been prepared by the EPO or JPO .....$930.00 -- International preliminary examination fee paid to USPTO (37 CFR 1.492).....$720.00 -- No international preliminary examination fee paid to USPTO (37 CFR 1.492) but international search fee paid to USPTO (37 CFR 1.445(a)(2)) .....$790.00 -- Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO.....$1,070.00 -- International preliminary examination fee paid to USPTO (37 CFR 1.482) and all claims satisfied provision of PCT Article 33(1) to (4).....$98.00							
ENTER APPROPRIATE BASIC FEE AMOUNT =						$ 930.00	
Surcharge of $130.00 for furnishing the National fee or oath or declaration later than ☐ 20 ☒ 30 mos. from the earliest claimed priority date (37 CFR 1.492(e)).						130.00	
CLAIMS	NUMBER FILED		NUMBER EXTRA	RATE			
Total Claims	**36**	-20 =	16	X	$22.00	$ 352.00	
Independent Claims	**5**	-3 =	2	X	$82.00	164.00	
Multiple Dependent Claims(s) (if applicable)				+$270.00		$ 270.00	
TOTAL OF ABOVE CALCULATIONS =						$ **1846.00**	
Reduction by ½ for filing by small entity, if applicable. Affidavit must be filed also. (Note 37 CFR 1.9, 1.27, 1.28).						0.00	
SUBTOTAL =						$ **1846.00**	
Processing fee of $130.00, for furnishing the English Translation later than ☐ 20 ☐ 30 mos., from the earliest claimed priority date (37 CFR 1.492(f).						0.00	
TOTAL NATIONAL FEE =						$ **1846.00**	
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31). $40.00 per property +						$ 0.00	
Fee for Petition to Revive Unintentionally Abandoned Application ($1,320 – Small Entity Fee = $660)						$ 0.00	
TOTAL FEES ENCLOSED =						$ **1846.00**	
						Amount to be refunded	$
						Charged	$

a. ☒ A check in the amount of $**1846.00** to cover the above fees is enclosed.
b. ☐ Please charge my Deposit Account No. 14-1140 in the amount of $_____ to cover the above fees. A duplicate copy of this form is enclosed.
c. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 14-1140. A duplicate copy of this form is enclosed.

SEND ALL CORRESPONDENCE TO:

NIXON & VANDERHYE P.C.
1100 North Glebe Road, 8th Floor
Arlington, Virginia 22201
Telephone: (703) 816-4000

Signature

Arthur R. Crawford
Name

25,327 — Registration Number
June 19, 1998 — Date

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of

HERMON-TAYLOR et al

Atty. Ref.: 117-260

Serial No. **(To Be Assigned)**

Group:

Filed: **19 June 1998**

Examiner:

For: **NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES IN PATHOGENIC MYCOBACTERIA AND THEIR USE AS DIAGNOSTICS, VACCINES AND TARGETS FOR CHEMOTHERAPY**

* * * * * * * * * * *

June 19, 1998

Honorable Commissioner of Patents
and Trademarks
Washington, DC 20231

Sir:

PRELIMINARY AMENDMENT

In order to place the above-identified application in better condition for examination, please amend the application as follows:

IN THE CLAIMS

Claim 4, lines 2 and 3, change "any one of claims 1 to 3" to -- Claim 1 or 2 --.

Claim 8, line 3, change "any one of claims 4 to 7" to -- Claim 4 --.

Claim 9, line 2, change "any one of claims 4 to 7" to -- Claim 4 --.

Claim 10, line 2, change "any one of claims 1 to 3" to -- Claim 1 or 2 --.

Please cancel claim 12 without prejudice.

Claim 13, lines 4 and 5, change "any one of claims 1 to 3" to -- Claim 1 or 2 --.

Claim 14, line 2, change "any one of claims 1 to 3" to -- Claim 1 or 2 --.

Claim 15, line 3, change "claims 1 to 3" to -- Claim 1 or 2 --.

Claim 16, line 2, change "any one of claims 1 to 3" to -- Claim 1 or 2 --.

Claim 18, line 3, change "claims 1 to 3" to -- Claim 1 or 2--.

Please delete Claim 19 without prejudice.

Claim 20, line 1, change "claims 18 or 19" to -- Claim 18 --.

Claim 21, lines 3 and 4, change "any one of claims 1 to 3" to -- Claim 1 or 2 --.

REMARKS

The above amendments are made to place the claims in a more traditional format.

Novel polynucleotides and polypeptides in pathogenic mycobacteria and their use as diagnostics, vaccines and targets for chemotherapy.

This invention relates to the novel polynucleotide sequence we have designated "GS" which we have identified in pathogenic mycobacteria. GS is a pathogenicity island within 8kb of DNA comprising a core region of 5.75kb and an adjacent transmissable element within 2.25kb. GS is contained within *Mycobacterium paratuberculosis, Mycobacterium avium* subsp. *silvaticum* and some pathogenic isolates of *M.avium*. Functional portions of the core region of GS are also represented by regions with a high degree of homology that we have identified in cosmids containing genomic DNA from *Mycobacterium tuberculosis*.

Background to the invention

Mycobacterium tuberculosis (Mtb) is a major cause of global diseases of humans as well as animals. Although conventional methods of diagnosis including microscopy, culture and skin testing exist for the recognition of these diseases, improved methods particularly new immunodiagnostics and DNA-based detection systems are needed. Drugs used to treat tuberculosis are increasingly encountering the problem of resistant organisms. New drugs targeted at specific pathogenicity determinants as well as new vaccines for the prevention and treatment of tuberculosis are required. The importance of *Mtb* as a global pathogen is reflected in the commitment being made to sequencing the entire genome of this organism. This has generated a large amount of DNA sequence data of genomic DNA within cosmid and other libraries. Although the DNA sequence is known in the art, the functions of the vast majority of these sequences, the proteins they encode, the biological significance of these proteins, and the overall relevance and use of these genes and their products as diagnostics, vaccines and targets for chemotherapy for tuberculous disease, remains entirely unknown.

Mycobacterium avium subsp.*silvaticum* (*Mavs*) is a pathogenic mycobacterium causing diseases of animals and birds, but it can

also affect humans. *Mycobacterium paratuberculosis* (*Mptb*) causes chronic inflammation of the intestine in many species of animals including primates and can also cause Crohn's disease in humans. *Mptb* is associated with other chronic inflammatory diseases of humans such as sarcoidosis. Subclinical *Mptb* infection is widespread in domestic livestock and is present in milk from infected animals. The organism is more resistant to pasteurisation than *Mtb* and can be conveyed to humans in retail milk supplies. *Mptb* is also present in water supplies, particularly those contaminated with run-off from heavily grazed pastures. *Mptb* and *Mavs* contain the insertion elements IS*900* and IS*902* respectively, and these are linked to pathogenicity in these organisms. IS*900* and IS*902* provide convenient highly specific multi-copy DNA targets for the sensitive detection of these organisms using DNA-based methods and for the diagnosis of infections in animals and humans. Much improvement is however required in the immunodiagnosis of *Mptb* and *Mavs* infections in animals and humans. *Mptb* and *Mavs* are in general, resistant in vivo to standard anti-tuberculous drugs. Although substantial clinical improvements in infections caused by *Mptb*, such as Crohn's disease, may result from treatment of patients with combinations of existing drugs such as Rifabutin, Clarithromycin or Azithromycin, additional effective drug treatments are required. Furthermore, there is an urgent need for effective vaccines for the prevention and treatment of *Mptb* and *Mavs* infections in animals and humans based upon the recognition of specific pathogenicity determinants.

Pathogenicity islands are, in general, 7-9kb regions of DNA comprising a core domain with multiple ORFs and an adjacent transmissable element. The transmissable element also encodes proteins which may be linked to pathogenicity, such as by providing receptors for cellular recognition. Pathogenicity islands are envisaged as mobile packages of DNA which, when they enter an organism, assist in bringing about its convertion from a non-disease-causing to a disease-causing strain.

Description of the Drawings

Figure 1(a) and (b) shows a linear map of the pathogenicity island GS in *Mavs* (Fig 1a) and in *Mptb* (Fig 1b). The main open reading frames are illustrated as ORFs A to H. ORFs A to F are found within the core region of GS. ORFs G and H are encoded by the adjacent transmissable element portion of GS.

Disclosure of the invention

Using a DNA-based differential analysis technology we have discovered and characterised a novel polynucleotide in *Mptb* (isolates 0022 from a Guernsey cow and 0021 from a red deer). This polynucleotide comprises the gene region we have designated GS. GS is found in *Mptb* using the identifier DNA sequences Seq.ID.No 1 and 2 where the Seq.ID No2 is the complementary sequence of Seq.ID No 1. GS is also identified in *Mavs*. The complete DNA sequence incorporating the positive strand of GS from an isolate of *Mavs* comprising 7995 nucleotides, including the core region of GS and adjacent transsmissable element, is given in Seq.ID No.3. DNA sequence comprising 4435 bp of the positive strand of GS obtained from an isolate of *Mptb* including the core region of GS (nucleotides 1614 to 6047 of GS in *Mavs*) is given in Seq.ID No 4. The DNA sequence of GS from *Mptb* is highly (99.4%) homologous to GS in *Mavs*. The remaining portion of the DNA sequence of GS in *Mptb*, is readily obtainable by a person skilled in the art using standard laboratory procedures. The entire functional DNA sequence including core region and transmisable element of GS in *Mptb* and *Mavs* as described above, comprise the polynucleotide sequences of the invention.

There are 8 open reading frames (ORFs) in GS. Six of these designated GSA, GSB, GSC, GSD, GSE and GSF are encoded by the core DNA region of GS which, characteristically for a pathogenicity island, has a different GC content than the rest of the microbial genome. Two ORFs designated GSG and GSH are encoded by the transmissable element of GS whose GC content resembles that of the rest of the mycobacterial genome. The ORF GSH comprises two sub-ORFs H_1 H_2 on the complementary DNA strand linked by a programmed frameshifting site so that a single polypeptide is translated from the ORF GSH. The nucleotide

sequences of the 8 ORFs in GS and their translations are shown in Seq. ID No 5 to Seq.ID No 29 as follows:

ORF A: Seq. ID No 5 Nucleotides 50 to 427 of GS from *Mavs*
Seq. ID No 6 Amino acid sequence encoded by Seq.ID No 5.

ORF B: Seq. ID No 7 Nucleotides 772 to 1605 of GS from *Mavs*
Seq. ID No 8 Amino acid sequence encoded by Seq.ID No 7.

ORF C: Seq. ID No 9 Nucleotides 1814 to 2845 of GS from *Mavs*
Seq. ID No 10 Amino acid sequence encoded by Seq.ID No 9.
Seq. ID No 11 Nucleotides 201 to 1232 of GS from *Mptb*
Seq. ID No 12 Amino acid sequence encoded by Seq.ID No 11

ORF D: Seq. ID No 13 Nucleotides 2785 to 3804 of GS from *Mavs*
Seq. ID No 14 Amino acid sequence encoded by Seq.ID No 13.
Seq. ID No 15 Nucleotides 1172 to 2191 of GS from *Mptb*
Seq. ID No 16 Amino acid sequence encoded by Seq.ID No 15.

ORF E: Seq. ID No 17 Nucleotides 4080 to 4802 of GS from *Mavs*
Seq. ID No 18 Amino acid sequence encoded by Seq.ID No 17.
Seq. ID No 19 Nucleotides 2467 to 3189 of GS from *Mptb*
Seq. ID No 20 Amino acid sequence encoded by Seq.ID No 19.

ORF F: Seq. ID No 21 Nucleotides 4947 to 5747 of GS from *Mavs*
Seq. ID No 22 Amino acid sequence encoded by Seq.ID No 21.
Seq. ID No 23 Nucleotides 3335 to 4135 of GS from *Mptb*
Seq. ID No 24 Amino acid sequence encoded by Seq.ID No 23.

ORF G: Seq. ID No 25 Nucleotides 6176 to 7042 of GS from *Mavs* Seq. ID No 26 Amino acid sequence encoded by Seq.ID No 25.

ORF H: Seq.ID No 27 Nucleotides 7953 to 6215 from *Mavs*.

ORF H_1: Seq.ID No 28 Amino acid sequence encoded by nucleotides 7953 to 7006 of Seq.ID No 27

ORF H_2: Seq.ID No 29 Amino acid sequence encoded by nucleotides 7009 to 6215 of Seq.ID No 27

The polynucleotides in *Mtb* with homology to the ORFs B, C, E and F of GS in *Mptb* and *Mavs*, and the polypeptides they are now known to encode as a result of our invention, are as follows:

ORF B: Seq.ID No 30 Cosmid MTCY277 nucleotides 35493 to 34705
Seq.ID No 31 Amino acid sequence encoded by Seq.ID No30.

ORF C: Seq.ID No 32 Cosmid MTCY277 nucleotides 31972 to 32994
Seq.ID No 33 Amino acid sequence encoded by Seq.ID No32.

ORF E: Seq.ID No 34 Cosmid MTCY277 nucleotides 34687 to 33956
Seq.ID No 35 Amino acid sequence encoded by Seq.ID No34.

ORF E: Seq.ID No 36 Cosmid MTO24 nucleotides 15934 to 15203
Seq.ID No 37 Amino acid sequence encoded by Seq.ID No36.

ORF F: Seq.ID No38 Cosmid MTO24 nucleotides 15133 to 14306
Seq.ID No 39 Amino acid sequence encoded by Seq.ID No38.

The proteins and peptides encoded by the ORFs A to H in *Mptb* and *Mavs* and the amino acid sequences from homologous genes we have

discovered in *Mtb* given in Seq.ID Nos 31, 33, 35, 37 and 39, as described above and fragments thereof, comprise the polypeptides of the invention. The polypeptides of the invention are believed to be associated with specific immunoreactivity and with the pathogenicity of the host micro-organisms from which they were obtained.

The present invention thus provides a polynucleotide in substantially isolated form which is capable of selectively hybridising to sequence ID Nos 3 or 4 or a fragment thereof. The polynucleotide fragment may alternatively comprise a sequence selected from the group of Seq.ID.No: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25 and 27. The invention further provides a polynucleotide in substantially isolated form whose sequence consists essentially of a sequence selected from the group Seq ID Nos. 30, 32, 34, 36 and 38, or a corresponding sequence selectively hybridizable thereto, or a fragment of said sequence or corresponding sequence.

The invention further provides diagnostic probes such as a probe which comprises a fragment of at least 15 nucleotides of a polynucleotide of the invention, or a peptide nucleic acid or similar synthetic sequence specific ligand, optionally carrying a revealing label. The invention also provides a vector carrying a polynucleotide as defined above, particularly an expression vector.

The invention further provides a polypeptide in substantially isolated form which comprises any one of the sequences selected from the group consisting Seq.ID.No: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 29, 31, 33, 35, 37 and 39, or a polypeptide substantially homologous thereto. The invention additionally provides a polypeptide fragment which comprises a fragment of a polypeptide defined above, said fragment comprising at least 10 amino acids and an epitope. The invention also provides polynucleotides in substantially isolated form which encode polypeptides of the invention, and vectors which comprise such polynucleotides, as well as antibodies capable of binding such polypeptides. In an additional aspect, the invention provides

kits comprising polynucleotides, polypeptides, antibodies or synthetic ligands of the invention and methods of using such kits in diagnosing the presence or absence of mycobacteria in a sample. The invention also provides pharmaceutical compositions comprising polynucleotides of the invention, polypeptides of the invention or antisense probes and the use of such compositions in the treatment or prevention of diseases caused by mycobacteria. The invention also provides polynucleotihe prevention and treatment of infections due to GS-containing pathogenic mycobacteria in animals and humans and as a means of enhacing in vivo susceptibility of said mycobacteria to antimicrobial drugs. The invention also provides bacteria or viruses transformed with polynucleotides of the invention for use as vaccines. The invention further provides *Mptb* or *Mavs* in which all or part or the polynucleotides of the invention have been deleted or disabled to provide mutated organisms of lower pathogenicity for use as vaccines in animals and humans. The invention further provides *Mtb* in which all or part of the polynucleotides encoding polypeptides of the invention have been deleted or disabled to provided mutated organisms or lower pathogenicity for use as vaccines in animals and humans.

A further aspect of the invention is our discovery of homologies between the ORFs B, C and E in GS on the one hand, and *Mtb* cosmid MTCY277 on the other (data from Genbank database using the computer programmes BLAST and BLIXEM). The homologous ORFs in MTCY277 are adjacent to one another consistent with the form of another pathogenicity island in *Mtb*. A further aspect of the invention is our discovery of homologies between ORFs E and F in GS, and *Mtb* cosmid MTO24 (also Genbank, as above) with the homologous ORFs close to one another. The use of polynucleotides and polypeptides from *Mtb* (Seq. ID Nos 30,31, 32, 33, 34, 35, 36, 37, 38 and 39) in substantially isolated form as diagnostics, vaccines and targets for chemotherapy, for the management and prevention of *Mtb* infections in humans and animals, and the processes involved in the preparation and use of these diagnostics, vaccines and new chemotherapeutic agents, comprise further aspects of the invention.

Detailed description of the invention.

A. Polynucleotides

Polynucleotides of the invention as defined herein may comprise DNA or RNA. They may also be polynucleotides which include within them synthetic or modified nucleotides or peptide nucleic acids. A number of different types of modification to oligonucleotides are known in the art. These include methylphosphonate and phosphorothioate backbones, addition of acridine or polylysine chains at the 3' and/or 5' ends of the molecule. For the purposes of the present invention, it is to be understood that the polynucleotides described herein may be modified by any method available in the art. Such modifications may be carried out in order to couple the said polynucleotide to a solid phase or to enhance the recognition, the *in vivo* activity, or the lifespan of polynucleotides of the invention.

A number of different types of polynucleotides of the invention are envisaged. In the broadest aspect, polynucleotides and fragments thereof capable of hybridizing to SEQ ID NO:3 or 4 form a first aspect of the invention. This includes the polynucleotide of SEQ ID NO: 3 or 4. Within this class of polynucleotides various sub-classes of polynucleotides are of particular interest.

One sub-class of polynucleotides which is of interest is the class of polynucleotides encoding the open reading frames A, B, C, D, E, F, G and H, including SEQ ID NOs:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25 and 27. As discussed below, polynucleotides encoding ORF H include the polynucleotide sequences 7953 to 7006 and 7009 to 6215 within SEQ ID NO: 27, as well as modified sequences in which the frame-shift has been modified so that the two sub-reading frames are placed in a single reading frame. This may be desirable where the polypeptide is to be produced in recombinant expression systems.

The invention thus provides a polynucleotide in substantially isolated form which encodes any one of these ORFs or combinations

thereof. Combinations thereof includes combinations of 2, 3, 4, 5 or all of the ORFs. Polynucleotides may be provided which comprise an individual ORF carried in a recombinant vector including the vectors described herein. Thus in one preferred aspect the invention provides a polynucleotide in substantially isolated form capable of selectively hybridizing to the nucleic acid comprising ORFs A to F of the core region of the *Mptb* and *Mavs* pathogenicity islands of the invention. Fragments thereof corresponding to ORFs A to E, B to F, A to D, B to E, A to C, B to D or any two adjacent ORFs are also included in the invention.

Polynucleotides of the invention will be capable of selectively hybridizing to the corresponding portion of the GS region, or to the corresponding ORFs of *Mtb* described herein. The term "selectively hybridizing" indicates that the polynucleotides will hybridize, under conditions of medium to high stringency (for example 0.03 M sodium chloride and 0.03 M sodium citrate at from about 50°C to about 60°C) to the corresponding portion of SEQ ID NO:3 or 4 or the complementary strands thereof but not to genomic DNA from mycobacteria which are usually non-pathogenic including non-pathogenic species of *M.avium*. Such polynucleotides will generally be generally at least 68%, e.g. at least 70%, preferably at least 80 or 90% and more preferably at least 95% homologous to the corresponding DNA of GS. The corresponding portion will be of over a region of at least 20, preferably at least 30, for instance at least 40, 60 or 100 or more contiguous nucleotides.

By "corresponding portion" it is meant a sequence from the GS region of the same or substantially similar size which has been determined, for example by computer alignment, to have the greatest degree of homology to the polynucleotide.

Any combination of the above mentioned degrees of homology and minimum sizes may be used to define polynucleotides of the invention, with the more stringent combinations (i.e. higher homology over longer lengths) being preferred. Thus for example a polynucleotide which is at least 80% homologous over 25, preferably 30 nucleotides forms one aspect of the invention, as

does a polynucleotide which is at least 90% homologous over 40 nucleotides.

A further class of polynucleotides of the invention is the class of polynucleotides encoding polypeptides of the invention, the polypeptides of the invention being defined in section B below. Due to the redundancy of the genetic code as such, polynucleotides may be of a lower degree of homology than required for selective hybridization to the GS region. However, when such polynucleotides encode polypeptides of the invention these polynucleotides form a further aspect. It may for example be desirable where polypeptides of the invention are produced recombinantly to increase the GC content of such polynucleotides. This increase in GC content may result in higher levels of expression via codon usage more appropriate to the host cell in which recombinant expression is taking place.

An additional class of polynucleotides of the invention are those obtainable from cosmids MTCY277 and MT024 (containing *Mtb* genomic sequences), which polynucleotides consist essentially of the fragment of the cosmid containing an open reading frame encoding any one of the homologous ORFs B, C, E or F respectively. Such polynucleotides are referred to below as *Mtb* polynucleotides. However, where reference is made to polynucleotides in general such reference includes *Mtb* polynucleotides unless the context is explicitly to the contrary. In addition, the invention provides polynucleotides which encode the same polypeptide as the abovementioned ORFs of *Mtb* but which, due to the redundancy of the genetic code, have different nucleotide sequences. These form further *Mtb* polynucleotides of the invention. Fragments of *Mtb* polynucleotides suitable for use as probes or primers also form a further aspect of the invention.

The invention further provides polynucleotides in substantially isolated form capable of selectively hybridizing (where selectively hybridizing is as defined above) to the *Mtb* polynucleotides of the invention.

The invention further provides the *Mtb* polynucleotides of the invention linked, at either the 5' and/or 3' end to polynucleotide sequences to which they are not naturally contiguous. Such sequences will typically be sequences found in cloning or expression vectors, such as promoters, 5' untranslated sequence, 3' untranslated sequence or termination sequences. The sequences may also include further coding sequences such as signal sequences used in recombinant production of proteins.

Further polynucleotides of the invention are illustrated in the accompanying examples.

Polynucleotides of the invention may be used to produce a primer, e.g. a PCR primer, a primer for an alternative amplification reaction, a probe e.g. labelled with a revealing label by conventional means using radioactive or non-radioactive labels or a probe linked covalently to a solid phase, or the polynucleotides may be cloned into vectors. Such primers, probes and other fragments will be at least 15, preferably at least 20, for example at least 25, 30 or 40 or more nucleotides in length, and are also encompassed by the term polynucleotides of the invention as used herein.

Primers of the invention which are preferred include primers directed to any part of the ORFs defined herein. The ORFs from other isolates of pathogenic mycobacteria which contain a GS region may be determined and conserved regions within each individual ORF may be identified. Primers directed to such conserved regions form a further preferred aspect of the invention. In addition, the primers and other polynucleotides of the invention may be used to identify, obtain and isolate ORFs capable of selectively hybridizing to the polynucleotides of the invention which are present in pathogenic mycobacteria but which are not part of a pathogenicity island in that particular species of bacteria. Thus in addition to the ORFs B, C, E and F which have been identified in *Mtb*, similar ORFs may be identified in other pathogens and ORFs corresponding to the GS ORFs C, D, E, F and H, may also be identified.

Polynucleotides such as DNA polynucleotides and probes according to the invention may be produced recombinantly, synthetically, or by any means available to those of skill in the art. They may also be cloned by standard techniques.

In general, primers will be produced by synthetic means, involving a step-wise manufacture of the desired nucleic acid sequence one nucleotide at a time. Techniques for accomplishing this using automated techniques are readily available in the art. Longer polynucleotides will generally be produced using recombinant means, for example using a PCR (polymerase chain reaction) cloning techniques. This will involve making a pair or primers (e.g. of about 15-30 nucleotides) to a region of GS, which it is desired to clone, bringing the primers into contact with genomic DNA from a mycobacterium or a vector carrying the GS sequence, performing a polymerase chain reaction under conditions which bring about amplification of the desired region, isolating the amplified fragment (e.g. by purifying the reaction mixture on an agarose gel) and recovering the amplified DNA. The primers may be designed to contain suitable restriction enzyme recognition sites so that the amplified DNA can be cloned into a suitable cloning vector.

Such techniques may be used to obtain all or part of the GS or ORF sequences described herein, as well as further genomic clones containing full open reading frames. Although in general such techniques are well known in the art, reference may be made in particular to Sambrook J., Fritsch EF., Maniatis T (1989). Molecular cloning: a Laboratory Manual, 2nd edn. Cold Spring Harbor, New York, Cold Spring Harbor Laboratory.

Polynucleotides which are not 100% homologous to the sequences of the present invention but fall within the scope of the invention can be obtained in a number of ways.

Other isolates or strains of pathogenic mycobacteria will be expected to contain allelic variants of the GS sequences described herein, and these may be obtained for example by probing genomic DNA libraries made from such isolates or strains

of bacteria using GS or ORF sequences as probes under conditions of medium to high stringency (for example 0.03M sodium chloride and 0.03M sodium citrate at from about 50°C to about 60°C).

A particularly preferred group of pathogenic mycobacteria are isolates of *M.paratuberculosis*. Polynucleotides based on GS regions from such bacteria are particularly preferred. Preferred fragments of such regions include fragments encoding individual open reading frames including the preferred groups and combinations of open reading frames discussed above.

Alternatively, such polynucleotides may be obtained by site directed mutagenesis of the GS or ORF sequences or allelic variants thereof. This may be useful where for example silent codon changes are required to sequences to optimise codon preferences for a particular host cell in which the polynucleotide sequences are being expressed. Other sequence changes may be desired in order to introduce restriction enzyme recognition sites, or to alter the property or function of the polypeptides encoded by the polynucleotides of the invention. Such altered property or function will include the addition of amino acid sequences of consensus signal peptides known in the art to effect transport and secretion of the modified polypeptide of the invention. Another altered property will include metagenesis of a catalytic residue or generation of fusion proteins with another polypeptide. Such fusion proteins may be with an enzyme, with an antibody or with a cytokine or other ligand for a receptor, to target a polypeptide of the invention to a specific cell type *in vitro* or *in vivo*.

The invention further provides double stranded polynucleotides comprising a polynucleotide of the invention and its complement.

Polynucleotides or primers of the invention may carry a revealing label. Suitable labels include radioisotopes such as ^{32}P or ^{35}S, enzyme labels, other protein labels or smaller labels such as biotin or fluorophores. Such labels may be added to polynucleotides or primers of the invention and may be detected using by techniques known per se.

Polynucleotides or primers of the invention or fragments thereof labelled or unlabelled may be used by a person skilled in the art in nucleic acid-based tests for the presence or absence of *Mptb*, *Mavs*, other GS-containing pathogenic mycobacteria, or *Mtb* applied to samples of body fluids, tissues, or excreta from animals and humans, as well as to food and environmental samples such as river or ground water and domestic water supplies.

Human and animal body fluids include sputum, blood, serum, plasma, saliva, milk, urine, csf, semen, faeces and infected discharges. Tissues include intestine, mouth ulcers, skin, lymph nodes, spleen, lung and liver obtained surgically or by a biopsy technique. Animals particularly include commercial livestock such as cattle, sheep, goats, deer, rabbits but wild animals and animals in zoos may also be tested.

Such tests comprise bringing a human or animal body fluid or tissue extract, or an extract of an environmental or food sample, into contact with a probe comprising a polynucleotide or primer of the invention under hybridising conditions and detecting any duplex formed between the probe and nucleic acid in the sample. Such detection may be achieved using techniques such as PCR or by immobilising the probe on a solid support, removing nucleic acid in the sample which is not hybridized to the probe, and then detecting nucleic acid which has hybridized to the probe. Alternatively, the sample nucleic acid may be immobilized on a solid support, and the amount of probe bound to such a support can be detected. Suitable assay methods of this any other formats can be found in for example WO89/03891 and WO90/13667.

Polynucleotides of the invention or fragments thereof labelled or unlabelled may also be used to identify and characterise different strains of *Mptb*, *Mavs*, other GS-containing pathogenic mycobacteria, or *Mtb*, and properties such as drug resistance or susceptibility.

The probes of the invention may conveniently be packaged in the form of a test kit in a suitable container. In such kits the probe may be bound to a solid support where the assay format for

which the kit is designed requires such binding. The kit may also contain suitable reagents for treating the sample to be probed, hybridising the probe to nucleic acid in the sample, control reagents, instructions, and the like.

The use of polynucleotides of the invention in the diagnosis of inflammatory diseases such as Crohn's disease or sarcoidosis in humans or Johne's disease in animals form a preferred aspect of the invention. The polynucleotides may also be used in the prognosis of these diseases. For example, the response of a human or animal subject in response to antibiotic, vaccination or other therapies may be monitored by utilizing the diagnostic methods of the invention over the course of a period of treatment and following such treatment.

The use of *Mtb* polynucleotides (particularly in the form of probes and primers) of the invention in the above-described methods form a further aspect of the invention, particularly for the detection, diagnosis or prognosis of *Mtb* infections.

B. Polypeptides.

Polypeptides of the invention include polypeptides in substantially isolated form encoded by GS. This includes the full length polypeptides encoded by the positive and complementary negative strands of GS. Each of the full length polypeptides will contain one of the amino acid sequences set out in Seq ID NOs:6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28 and 29. Polypeptides of the invention further include variants of such sequences, including naturally occurring allelic variants and synthetic variants which are substantially homologous to said polypeptides. In this context, substantial homology is regarded as a sequence which has at least 70%, e.g. 80%, 90%, 95% or 98% amino acid homology (identity) over 30 or more, e.g 40, 50 or 100 amino acids. For example, one group of substantially homolgous polypeptides are those which have at least 95% amino acid identity to a polypeptide of any one of Seq ID NOs:6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28 and 29 over their entire length. Even more preferably, this homology is 98%.

Polypeptides of the invention further include the polypeptide sequences of the homologous ORFs of *Mtb*, namely Seq ID Nos. 31, 33, 35, 37 and 39. Unless explicitly specified to the contrary, reference to polypeptides of the invention and their fragments include these *Mtb* polypeptides and fragments, and variants thereof (substanially homologous to said sequences) as defined herein.

Polypeptides of the invention may be obtained by the standard techniques mentioned above. Polypeptides of the invention also include fragments of the above mentioned full length polypeptides and variants thereof, including fragments of the sequences set out in SEQ ID NOs:6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 29, 31, 33, 35, 37 and 39. Such fragments for example of 8, 10, 12, 15 or up to 30 or 40 amino acids may also be obtained synthetically using standard techniques known in the art.

Preferred fragments include those which include an epitope, especially an epitope which is specific to the pathogenicity of the mycobacterial cell from which the polypeptide is derived. Suitable fragments will be at least about 5, e.g. 8, 10, 12, 15 or 20 amino acids in size, or larger. Epitopes may be determined either by techniques such as peptide scanning techniques as described by Geysen et al, Mol.Immunol., 23; 709-715 (1986), as well as other techniques known in the art.

The term "an epitope which is specific to the pathogenicity of the mycobacterial cell" means that the epitope is encoded by a portion of the GS region, or by the corresponding ORF sequences of *Mtb* which can be used to distinguish mycobacteria which are pathogenic by from related non-pathogenic mycobacteria including non-pathogenic species of *M.avium*. This may be determined using routine methodology. A candidate epitope from an ORF may be prepared and used to immunise an animal such as a rat or rabbit in order to generate antibodies. The antibodies may then be used to detect the presence of the epitope in pathogenic mycobacteria and to confirm that non-pathogenic mycobacteria do not contain any proteins which react with the epitope. Epitopes may be linear or conformational.

Polypeptides of the invention may be in a substantially isolated form. It will be understood that the polypeptide may be mixed with carriers or diluents which will not interfere with the intended purpose of the polypeptide and still be regarded as substantially isolated. A polypeptide of the invention may also be in a substantially purified form, in which case it will generally comprise the polypeptide in a preparation in which more than 90%, e.g. 95%, 98% or 99% of the polypeptide in the preparation is a polypeptide of the invention.

Polypeptides of the invention may be modified to confer a desired property or function for example by the addition of Histidine residues to assist their purification or by the addition of a signal sequence to promote their secretion from a cell.

Thus, polypeptides of the invention include fusion proteins which comprise a polypeptide encoding all or part of one or more of an ORF of the invention fused at the N- or C-terminus to a second sequence to provide the desired property or function. Sequences which promote secretion from a cell include, for example the yeast α-factor signal sequence.

A polypeptide of the invention may be labelled with a revealing label. The revealing label may be any suitable label which allows the polypeptide to be detected. Suitable labels include radioisotopes, e.g. ^{125}I, ^{35}S enzymes, antibodies, polynucleotides and ligands such as biotin. Labelled polypeptides of the invention may be used in diagnostic procedures such as immunoassays in order to determine the amount of a polypeptide of the invention in a sample. Polypeptides or labelled polypeptides of the invention may also be used in serological or cell mediated immune assays for the detection of immune reactivity to said polypeptides in animals and humans using standard protocols.

A polypeptide or labelled polypeptide of the invention or fragment thereof may also be fixed to a solid phase, for example the surface of an immunoassay well, microparticle, dipstick or biosensor. Such labelled and/or immobilized polypeptides may be

packaged into kits in a suitable container along with suitable reagents, controls, instructions and the like.

Such polypeptides and kits may be used in methods of detection of antibodies or cell mediated immunoreactivity, to the mycobacterial proteins and peptides encoded by the ORFs of the invention and their allelic variants and fragments, using immunoassay. Such host antibodies or cell mediated immune reactivity will occur in humans or animals with an immune system which detects and reacts against polypeptides of the invention. The antibodies may be present in a biological sample from such humans or animals, where the biological sample may be a sample as defined above particularly blood, milk or saliva.

Immunoassay methods are well known in the art and will generally comprise:

(a) providing a polypeptide of the invention comprising an epitope bindable by an antibody against said mycobacterial polypeptide;

(b) incubating a biological sample with said polypeptide under conditions which allow for the formation of an antibody-antigen complex; and

(c) determining whether antibody-antigen complex comprising said polypeptide is formed.

Immunoassay methods for cell mediated immune reactivity in animals and humans are also well known in the art (e.g. as described by Weir et al 1994, J.Immunol Methods 176; 93-101) and will generally comprise

(a) providing a polypeptide of the invention comprising an epitope bindable by a lymphocyte or macrophage or other cell receptor;

(b) incubating a cell sample with said polypeptide under conditions which allow for a cellular immune response such as release of cytokines or other mediator to occur; and

(c) detecting the presence of said cytokine or mediator in the incubate.

Polypeptides of the invention may be made by standard synthetic means well known in the art or recombinantly, as described below.

Polypeptides of the invention or fragments thereof labelled or unlabelled may also be used to identify and characterise different strains of *Mptb*, *Mavs*, other GS-containing pathogenic mycobacteria, or *Mtb*, and properties such as drug resistance or susceptibility.

The polypeptides of the invention may conveniently be packaged in the form of a test kit in a suitable container. In such kits the polypeptide may be bound to a solid support where the assay format for which the kit is designed requires such binding. The kit may also contain suitable reagents for treating the sample to be examined, control reagents, instructions, and the like.

The use of polypeptides of the invention in the diagnosis of inflammatory diseases such as Crohn's disease or sarcoidosis in humans or Johne's disease in animals form a preferred aspect of the invention. The polypeptides may also be used in the prognosis of these diseases. For example, the response of a human or animal subject in response to antibiotic or other therapies may be monitored by utilizing the diagnostic methods of the invention over the course of a period of treatment and following such treatment.

The use of *Mtb* polypeptides of the invention in the above-described methods form a further aspect of the invention, particularly for the detection, diagnosis or prognosis of *Mtb* infections.

Polypeptides of the invention may also be used in assay methods for identifying candidate chemical compounds which will be useful in inhibiting, binding to or disrupting the function of said polypeptides required for pathogenicity. In general, such assays involve bringing the polypeptide into contact with a candidate inhibitor compound and observing the ability of the compound to disrupt, bind to or interfer with the polypeptide.

There are a number of ways in which the assay may be formatted. For example, those polypeptides which have an enzymatic function may be assayed using labelled substrates for the enzyme, and the amount of, or rate of, conversion of the substrate into a product measured, e.g by chromatograpy such as HPLC or by a colourimetric assay. Suitable labels include ^{35}S, ^{125}I, biotin or enzymes such as horse radish peroxidase.

For example, the gene product of ORF C is believed to have GDP-mannose dehydratase activty. Thus an assay for inhbitors of the gene product may utilise for example labelled GDP-mannose, GDP or mannose and the activity of the gene product followed. ORF D encodes a gene related to the synthesis and regulation of capuslar polysaccharides, which are often associated with invasiveness and pathogenicity. Labelled polysaccharide substrates may be used in assays of the ORF D gene product. The gene product of ORF F encodes a protein with putative glucosyl transferase activity and thus labelled amino sugars such as β-1-3-N-acetylglucosamine may be used as substrates in assays.

Candidate chemical compounds which may be used may be natural or synthetic chemical compounds used in drug screening programmes. Extracts of plants which contain several characterised or uncharacterised components may also be used.

Alternatively, the a polypeptide of the invention may be screened against a panel of peptides, nucleic acids or other chemical functionalities which are generated by combinatorial chemistry. This will allow the definition of chemical entities which bind to polypeptides of the invention. Typically, the polypeptide of the invention will be brought into contact with a panel of compounds from a combinantorial library, with either the panel or the polypeptide being immobilized on a solid phase, under conditions suitable for the polypeptide to bind to the panel. The solid phase will then be washed under conditions in which only specific interactions between the polypeptide and individual members of the panel are retained, and those specific members may be utilized in further assays or used to design further panels of candidate compounds.

For example, a number of assay methods to define peptide interaction with peptides are known. For example, WO86/00991 describes a method for determining mimotopes which comprises making panels of catamer preparations, for example octamers of amino acids, at which one or more of the positions is defined and the remaining positions are randomly made up of other amino acids, determining which catamer binds to a protein of interest and re-screening the protein of interest against a further panel based on the most reactive catamer in which one or more additional designated positions are systematically varied. This may be repeated throughout a number of cycles and used to build up a sequence of a binding candidate compound of interest.

WO89/03430 describes screening methods which permit the preparation of specific mimotopes which mimic the immunological activity of a desired analyte. These mimotopes are identified by reacting a panel of individual peptides wherein said peptides are of systematically varying hydrophobicity, amphipathic characteristics and charge patterns, using an antibody against an antigen of interest. Thus in the present case antibodies against the a polypeptide of the inventoin may be employed and mimotope peptides from such panels may be identified.

C. Vectors.

Polynucleotides of the invention can be incorporated into a recombinant replicable vector. The vector may be used to replicate the nucleic acid in a compatible host cell. Thus in a further embodiment, the invention provides a method of making polynucleotides of the invention by introducing a polynucleotide of the invention into a replicable vector, introducing the vector into a compatible host cell, and growing the host cell under conditions which bring about replication of the vector. The vector may be recovered from the host cell. Suitable host cells are described below in connection with expression vectors.

D. Expression Vectors.

Preferably, a polynucleotide of the invention in a vector is operably linked to a control sequence which is capable of providing for the expression of the coding sequence by the host cell, i.e. the vector is an expression vector. The term "operably linked" refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences. Such vectors may be transformed into a suitable host cell as described above to provide for expression of a polypeptide of the invention. Thus, in a further aspect the invention provides a process for preparing polypeptides according to the invention which comprises cultivating a host cell transformed or transfected with an expression vector as described above, under conditions to provide for expression by the vector of a coding sequence encoding the polypeptides, and recovering the expressed polypeptides.

A further embodiment of the invention provides vectors for the replication and expression of polynucleotides of the invention, or fragments thereof. The vectors may be for example, plasmid, virus or phage vectors provided with an origin of replication, optionally a promoter for the expression of the said polynucleotide and optionally a regulator of the promoter. The vectors may contain one or more selectable marker genes, for example an ampicillin resistance gene in the case of a bacterial plasmid or a neomycin resistance gene for a mammalian vector. Vectors may be used *in vitro*, for example for the production of RNA or used to transfect or transform a host cell. The vector may also be adapted to be used *in vivo*, for example in a method of naked DNA vaccination or gene therapy. A further embodiment of the invention provides host cells transformed or transfected with the vectors for the replication and expression of polynucleotides of the invention, including the DNA of GS, the open reading frames thereof and other corresponding ORFs particularly ORFs B, C, E and F from *Mtb*. The cells will be chosen to be compatible with the said vector and may for example be bacterial, yeast, insect or mammalian.

Expression vectors are widely available in the art and can be obtained commercially. Mammalian expression vectors may comprise a mammalian or viral promoter. Mammalian promoters include the metallothionien promoter. Viral promoters include promoters from adenovirus, the SV40 large T promoter and retroviral LTR promoters. Promoters compatible with insect cells include the polyhedrin promoter. Yeast promoters include the alcohol dehydrogenase promoter. Bacterial promoters include the β-galactosidase promoter.

The expression vectors may also comprise enhancers, and in the case of eukaryotic vectors polyadenylation signal sequence downstream of the coding sequence being expressed.

Polypeptides of the invention may be expressed in suitable host cells, for example bacterial, yeast, plant, insect and mammalian cells, and recovered using standard purification techniques including, for example affinity chromatography, HPLC or other chromatographic separation techniques.

Polynucleotides according to the invention may also be inserted into the vectors described above in an antisense orientation in order to provide for the production of antisense RNA. Antisense RNA or other antisense polynucleotides or ligands may also be produced by synthetic means. Such antisense polynucleotides may be used in a method of controlling the levels of the proteins encoded by the ORFs of the invention in a mycobacterial cell.

Polynucleotides of the invention may also be carried by vectors suitable for gene therapy methods. Such gene therapy methods include those designed to provide vaccination against diseases caused by pathogenic mycobacteria or to boost the immune response of a human or animal infected with a pathogenic mycobacteria.

For example, Ziegner *et al*, AIDS, 1995, 9;43-50 describes the use of a replication defective recombinant amphotropic retrovirus to boost the immune response in patients with HIV infection. Such a retrovirus may be modified to carry a polynucleotide encoding a polypeptide or fragment thereof of the invention and the

retrovirus delivered to the cells of a human or animal subject in order to provide an immune response against said polypeptide. The retrovirus may be delivered directly to the patient or may be used to infecte cells *ex-vivo*, e.g. fibroblast cells, which are then introduced into the patient, optionally after being inactivated. The cells are desirably autologous or HLA-matched cells from the human or animal subject.

Gene therapy methods including methods for boosting an immune response to a particluar pathogen are disclosed generally in for example WO95/14091, the disclosure of which is incoporated herein by reference. Recombinant viral vectors include retroviral vectors, adenoviral vectors, adeno-associated viral vectors, vaccinia virus vectors, herpes virus vectors and alphavirus vectors. Alpha virus vectors are described in, for example, WO95/07994, the disclosure of which is incorporated herein by reference.

Where direct administration of the recombinant viral vector is contemplated, either in the form of naked nucleic acid or in the form of packaged particles carrying the nucleic acid this may be done by any suitable means, for example oral administration or intravenous injection. From 10^5 to 10^8 c.f.u of virus represents a typical dose, which may be repeated for example weekly over a period of a few months. Administration of autologous or HLA-matched cells infected with the virus may be more convenient in some cases. This will generally be achieved by administering doses, for example from 10^5 to 10^8 cells per dose which may be repeated as described above.

The recombinant viral vector may further comprise nucleic acid capable of expressing an accessory molecule of the immune system designed to increase the immune response. Such a moleclue may be for example and interferon, particularly interferon gamma, an interleukin, for example IL-1α, IL-1β or IL-2, or an HLA class I or II moleclue. This may be particularly desirable where the vector is intended for use in the treatment of humans or animals already infected with a mycobacteria and it is desired to boost the immune response.

E. Antibodies.

The invention also provides monoclonal or polyclonal antibodies to polypeptides of the invention or fragments thereof. The invention further provides a process for the production of monoclonal or polyclonal antibodies to polypeptides of the invention. Monoclonal antibodies may be prepared by conventional hybridoma technology using the polypeptides of the invention or peptide fragments thereof, as immunogens. Polyclonal antibodies may also be prepared by conventional means which comprise inoculating a host animal, for example a rat or a rabbit, with a polypeptide of the invention or peptide fragment thereof and recovering immune serum.

In order that such antibodies may be made, the invention also provides polypeptides of the invention or fragments thereof haptenised to another polypeptide for use as immunogens in animals or humans.

For the purposes of this invention, the term "antibody", unless specified to the contrary, includes fragments of whole antibodies which retain their binding activity for a polypeptide of the invention. Such fragments include Fv, F(ab') and $F(ab')_2$ fragments, as well as single chain antibodies. Furthermore, the antibodies and fragments thereof may be humanised antibodies, e.g. as described in EP-A-239400.

Antibodies may be used in methods of detecting polypeptides of the invention present in biological samples (where such samples include the human or animal body samples, and environmental samples, mentioned above) by a method which comprises:

(a) providing an antibody of the invention;

(b) incubating a biological sample with said antibody under conditions which allow for the formation of an antibody-antigen complex; and

(c) determining whether antibody-antigen complex comprising said antibody is formed.

Antibodies of the invention may be bound to a solid support for example an immunoassay well, microparticle, dipstick or biosensor and/or packaged into kits in a suitable container along with suitable reagents, controls, instructions and the like.

Antibodies of the invention may be used in the detection, diagnosis and prognosis of diseases as described above in relation to polypeptides of the invention.

F. Compositions.

The present invention also provides compositions comprising a polynucleotide or polypeptide of the invention together with a carrier or diluent. Compositions of the invention also include compositions comprising a nucleic acid, particularly and expression vector, of the invention. Compositions further include those carrying a recombinant virus of the invention. Such compositions include pharmaceutical compositions in which case the carrier or diluent will be pharmaceutically acceptable.

Pharmaceutically acceptable carriers or diluents include those used in formulations suitable for inhalation as well as oral, parenteral (e.g. intramuscular or intravenous or transcutaneous) administration. The formulations may conveniently be presented in unit dosage form and may be prepared by any of the methods well known in the art of pharmacy. Such methods include the step of bringing into association the active ingredient with the carrier which constitutes one or more accessory ingredients. In general the formulations are prepared by uniformly and intimately bringing into association the active ingredient with liquid carriers or finely divided solid carriers or both, and then, if necessary, shaping the product.

For example, formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the blood of the intended recipient, and aqueous and non-aqueous sterile suspensions which may include suspending agents and thickening

agents, and liposomes or other microparticulate systems which are designed to target the polynucleotide or the polypeptide of the invention to blood components or one or more organs, or to target cells such as M cells of the intestine after oral administration.

G. Vaccines.

In another aspect, the invention provides novel vaccines for the prevention and treatment of infections caused by *Mptb*, *Mavs*, other GS-containing pathogenic mycobacteria and *Mtb* in animals and humans. The term "vaccine" as used herein means an agent used to stimulate the immune system of a vertebrate, particularly a warm blooded vertebrate including humans, so as to provide protection against future harm by an organism to which the vaccine is directed or to assist in the eradication of an organism in the treatment of established infection. The immune system will be stimulated by the production of cellular immunity antibodies, desirably neutralizing antibodies, directed to epitopes found on or in a pathogenic mycobacterium which expresses any one of the ORFs of the invention. The antibody so produced may be any of the immunological classes, such as the immunoglobulins A, D, E, G or M. Vaccines which stimulate the production of IgA are interest since this is the principle immunoglobulin produced by the secretory system of warm-blooded animals, and the production of such antibodies will help prevent infection or colonization of the intestinal tract. However an IgM and IgG response will also be desirable for systemic infections such as Crohn's disease or tuberculosis.

Vaccines of the invention include polynucleotides of the invention or fragments thereof in suitable vectors and administered by injection of naked DNA using standard protocols. Polynucleotides of the invention or fragments thereof in suitable vectors for the expression of the polypeptides of the invention may be given by injection, inhalation or by mouth. Suitable vectors include *M.bovis* BCG, *M.smegmatis* or other mycobacteria, *Corynebacteria*, *Salmonella* or other agents according to established protocols.

Polypeptides of the invention or fragments thereof in substantially isolated form may be used as vaccines by injection, inhalation, oral administration or by transcutaneous application according to standard protocols. Adjuvants (such as Iscoms or polylactide-coglycolide encapsulation), cytokines such as IL-12 and other immunomodulators may be used for the selective enhancement of the cell mediated or humoral immunological responses. Vaccination with polynucleotides and/or polypeptides of the invention may be undertaken to increase the susceptibility of pathogenic mycobacteria to antimicrobial agents *in vivo*.

In instances wherein the polypeptide is correctly configured so as to provide the correct epitope, but is too small to be immunogenic, the polypeptide may be linked to a suitable carrier.

A number of techniques for obtaining such linkage are known in the art, including the formation of disulfide linkages using N-succinimidyl-3-(2-pyridylthio) propionate (SPDP) and succinimidyl 4-(N-maleimido-methyl)cyclohexane-1-carboxylate (SMCC) obtained from Pierce Company, Rockford, Illinois, (if the peptide lacks a sulfhydryl group, this can be provided by addition of a cysteine residue). These reagents create a disulfide linkage between themselves and peptide cysteine residues on one protein and an amide linkage through the epsilon-amino on a lysine, or other free amino group in the other. A variety of such disulfide/amide-forming agents are known. See, for example, Immun Rev (1982) 62:185. Other bifunctional coupling agents form a thioether rather than a disulfide linkage. Many of these thio-ether-forming agents are commercially available and include reactive esters of 6-maleimidocaproic acid, 2-bromoacetic acid, 2-iodoacetic acid, 4-(N-maleimido-methyl)cyclohexane-1-carboxylic acid, and the like. The carboxyl group can be activated by combining them with succinimide or 1-hydroxyl-2-nitro-4-sulfonic acid, sodium salt. Additional methods of coupling antigens employs the rotavirus/"binding peptide" system described in EPO Pub. No. 259,149, the disclosure of which is incorporated herein by reference. The foregoing list is not meant to be exhaustive, and modifications of the named compounds can clearly be used.

Any carrier may be used which does not itself induce the production of antibodies harmful to the host. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins; polysaccharides, such as latex functionalized Sepharose®, agarose, cellulose, cellulose beads and the like; polymeric amino acids, such as polyglutamic acid, polylysine, polylactide-coglycolide and the like; amino acid copolymers; and inactive virus particles. Especially useful protein substrates are serum albumins, keyhole limpet hemocyanin, immunoglobulin molecules, thyroglobulin, ovalbumin, tetanus toxoid, and other proteins well known to those skilled in the art.

The immunogenicity of the epitopes may also be enhanced by preparing them in mammalian or yeast systems fused with or assembled with particle-forming proteins such as, for example, that associated with hepatitis B surface antigen. See, e.g., US-A-4,722,840. Constructs wherein the epitope is linked directly to the particle-forming protein coding sequences produce hybrids which are immunogenic with respect to the epitope. In addition, all of the vectors prepared include epitopes specific to HBV, having various degrees of immunogenicity, such as, for example, the pre-S peptide.

In addition, portions of the particle-forming protein coding sequence may be replaced with codons encoding an epitope of the invention. In this replacement, regions which are not required to mediate the aggregation of the units to form immunogenic particles in yeast or mammals can be deleted, thus eliminating additional HBV antigenic sites from competition with the epitope of the invention.

Vaccines may be prepared from one or more immunogenic polypeptides of the invention. These polypeptides may be expressed in various host cells (e.g., bacteria, yeast, insect, or mammalian cells), or alternatively may be isolated from viral preparations or made synthetically.

In addition to the above, it is also possible to prepare live vaccines of attenuated microorganisms which express one or more

recombinant polypeptides of the invention. Suitable attenuated microorganisms are known in the art and include, for example, viruses (e.g., vaccinia virus), as well as bacteria.

The preparation of vaccines which contain an immunogenic polypeptide(s) as active ingredients, is known to one skilled in the art. Typically, such vaccines are prepared as injectables, or as suitably encapsulated oral preparations and either liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid prior to injestion or injection may also be prepared. The preparation may also be emulsified, or the protein encapsulated in liposomes. The active immunogenic ingredients are often mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients are, for example, water, saline, dextrose, glycerol, ethanol, or the like and combinations thereof. In addition, if desired, the vaccine may contain minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents, and/or adjuvants which enhance the effectiveness of the vaccine. Examples of adjuvants which may be effective include but are not limited to: aluminum hydroxide, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-nor-muramyl-L-alanyl-D-isoglutamine (CGP 11637, referred to as nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (CGP 19835A, referred to as MTP-PE), and RIBI, which contains three components extracted from bacteria, monophosphoryl lipid A, trehalose dimycolate and cell wall skeleton (MPL+TDM+CWS) in a 2% squalene/Tween® 80 emulsion. The effectiveness of an adjuvant may be determined by measuring the amount of antibodies directed against an immunogenic polypeptide containing an antigenic sequence resulting from administration of this polypeptide in vaccines which are also comprised of the various adjuvants.

The vaccines are conventionally administered parenterally, by injection, for example, either subcutaneously or intramuscularly. Additional formulations which are suitable for other modes of administration include suppositories, oral formulations or as

enemas. For suppositories, traditional binders and carriers may include, for example, polyalkylene glycols or triglycerides; such suppositories may be formed from mixtures containing the active ingredient in the range of 0.5% to 10%, preferably 1% - 2%. Oral formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. These compositions take the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders and contain 10% - 95% of active ingredient, preferably 25% - 70%.

The proteins may be formulated into the vaccine as neutral or salt forms. Pharmaceutically acceptable salts include the acid addition salts (formed with free amino groups of the peptide) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids such as acetic, oxalic, tartaric, maleic, and the like. Salts formed with the free carboxyl groups may also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine, procaine, and the like.

The vaccines are administered in a manner compatible with the dosage formulation, and in such amount as will be prophylactically and/or therapeutically effective. The quantity to be administered, which is generally in the range of 5μg to 250μg, of antigen per dose, depends on the subject to be treated, capacity of the subject's immune system to synthesize antibodies, mode of administration and the degree of protection desired. Precise amounts of active ingredient required to be administered may depend on the judgement of the practitioner and may be peculiar to each subject.

The vaccine may be given in a single dose schedule, or preferably in a multiple dose schedule. A multiple dose schedule is one in which a primary course of vaccination may be with 1-10 separate doses, followed by other doses given at subsequent time intervals

required to maintain and or reenforce the immune response, for example, at 1-4 months for a second dose, and if needed, a subsequent dose(s) after several months. The dosage regimen will also, at least in part, be determined by the need of the individual and be dependent upon the judgement of the practitioner.

In a further aspect of the invention, there is provided an attenuated vaccine comprising a normally pathogenic mycobacteria which harbours an attenuating mutation in any one of the genes encoding a polypeptide of the invention. The gene is selected from the group of ORFs A, B, C, D, E, F, G and H, including the homologous ORFs B, C, E and F in *Mtb*.

The mycobacteria may be used in the form of killed bacteria or as a live attenuated vaccine. There are advantages to a live attenuated vaccine. The whole live organism is used, rather than dead cells or selected cell components which may exhibit modified or denatured antigens. Protein antigens in the outer membrane will maintain their tertiary and quaternary structures. Therefore the potential to elicit a good protective long term immunity should be higher.

The term "mutation" and the like refers to a genetic lesion in a gene which renders the gene non-functional. This may be at either the level of transcription or translation. The term thus envisages deletion of the entire gene or substantial portions thereof, and also point mutations in the coding sequence which result in truncated gene products unable to carry out the normal function of the gene.

A mutation introduced into a bacterium of the invention will generally be a non-reverting attenuating mutation. Non-reverting means that for practical purposes the probability of the mutated gene being restored to its normal function is small, for example less than 1 in 10^6 such as less than 1 in 10^9 or even less than 1 in 10^{12}.

An attenuated mycobacteria of the invention may be in isolated form. This is usually desirable when the bacterium is to be used for the purposes of vaccination. The term "isolated" means that the bacterium is in a form in which it can be cultured, processed or otherwise used in a form in which it can be readily identified and in which it is substantially uncontaminated by other bacterial strains, for example non-attenuated parent strains or unrelated bacterial strains. The term "isolated bacterium" thus encompasses cultures of a bacterial mutant of the invention, for example in the form of colonies on a solid medium or in the form of a liquid culture, as well as frozen or dried preparations of the strains.

In a preferred aspect, the attenuated mycobacterium further comprises at least one additional mutation. This may be a mutation in a gene responsible for the production of products essential to bacterial growth which are absent in a human or animal host. For example, mutations to the gene for aspartate semi-aldehyde dehydrogenase (*asd*) have been proposed for the production of attenuated strains of Salmonella. The *asd* gene is described further in Gene (1993) 129; 123-128. A lesion in the *asd* gene, encoding the enzyme aspartate β-semialdehyde dehydrogenase would render the organism auxotrophic for the essential nutrient diaminopelic acid (DAP), which can be provided exogenously during bulk culture of the vaccine strain. Since this compound is an essential constituent of the cell wall for gram-negative and some gram-positive organisms and is absent from mammalian or other vertebrate tissues, mutants would undergo lysis after about three rounds of division in such tissues. Analogous mutations may be made to the attenuated mycobacteria of the invention.

In addition or in the alternative, the attenuated mycobacteria may carry a *recA* mutation. The *recA* mutation knocks out homologous recombination - the process which is exploited for the construction of the mutations. Once the *recA* mutation has been incorporated the strain will be unable to repair the constructed deletion mutations. Such a mutation will provide attenuated strains in which the possibility of homologous recombination to

with DNA from wild-type strains has been minimized. *RecA* genes have been widely studied in the art and their sequences are available. Further modifications may be made for additional safety.

The invention further provides a process for preparing a vaccine composition comprising an attenuated bacterium according to the invention process comprises (a) inoculating a culture vessel containing a nutrient medium suitable for growth of said bacterium; (b) culturing said bacterium; (c) recovering said bacteria and (d) mixing said bacteria with a pharmaceutically acceptable diluent or carrier.

Attenuated bacterial strains according to the invention may be constructed using recombinant DNA methodology which is known per se. In general, bacterial genes may be mutated by a process of targeted homologous recombination in which a DNA construct containing a mutated form of the gene is introduced into a host bacterium which it is desired to attenuate. The construct will recombine with the wild-type gene carried by the host and thus the mutated gene may be incorporated into the host genome to provide a bacterium of the present invention which may then be isolated.

The mutated gene may be obtained by introducing deletions into the gene, e.g by digesting with a restriction enzyme which cuts the coding sequence twice to excise a portion of the gene and then religating under conditions in which the excised portion is not reintroduced into the cut gene. Alternatively frame shift mutations may be introduced by cutting with a restriction enzyme which leaves overhanging 5' and 3' termini, filling in and/or trimming back the overhangs, and religating. Similar mutations may be made by site directed mutagenesis. These are only examples of the types of techniques which will readily be at the disposal of those of skill in the art.

Various assays are available to detect successful recombination. In the case of attenuations which mutate a target gene necessary for the production of an essential metabolite or catabolite

compound, selection may be carried out by screening for bacteria unable to grow in the absence of such a compound. Bacteria may also be screened with antibodies or nucleic acids of the invention to determine the absence of production of a mutated gene product of the invention or to confirm that the genetic lesion introduced - e.g. a deletion - has been incorporated into the genome of the attenuated strain.

The concentration of the attenuated strain in the vaccine will be formulated to allow convenient unit dosage forms to be prepared. Concentrations of from about 10^4 to 10^9 bacteria per ml will generally be suitable, e.g. from about 10^5 to 10^8 such as about 10^6 per ml. Live attenuated organisms may be administered subcutaneously or intramuscularly at up to 10^8 organisms in one or more doses, e.g from around 10^5 to 10^8, e.g about 10^6 or 10^7 organisms in a single dose.

The vaccines of the invention may be administered to recipients to treat established disease or in order to protect them against diseases caused by the corresponding wild type mycobacteria, such as inflammatory diseases such as Crohn's disease or sarcoidosis in humans or Johne's disease in animals. The vaccine may be administered by any suitable route. In general, subcutaneous or intramuscular injection is most convenient, but oral, intranasal and colorectal administration may also be used.

The following Examples illustrates aspects of the invention.

EXAMPLE 1

Tests for the presence of the GS identifier sequence were performed on 5μl bacterial DNA extracts (25 μg/ml to 500 μg/ml) using polymerase chain reaction based on the oligonucleotide primers 5'-GATGCCGTGAGGAGGTAAAGCTGC-3' (Seq ID No. 40) and 5'-GATACGGCTCTTGAATCCTGCACG-3' (Seq ID No. 41) from within the identifier DNA sequences (Seq.ID Nos 1 and 2). PCR was performed for 40 cycles in the presence of 1.5 mM magnesium and an annealing temperature of 58°C. The presence or absence of the correct amplification product indicated the presence or absence

of GS identifier sequence in the corresponding bacterium. GS identifier sequence is shown to be present in all the laboratory and field strains of *Mptb* and *Mavs* tested. This includes *Mptb* isolates 0025 (bovine CVL Weybridge), 0021 (caprine, Moredun), 0022 (bovine, Moredun), 0139 (human, Chiodini 1984), 0209, 0208, 0211, 0210, 0212, 0207, 0204, 0206 (bovine, Whipple 1990). All *Mptb* strains were IS*900* positive. The *Mavs* strains include 0010 and 0012 (woodpigeon, Thorel) 0018 (armadillo, Portaels) and 0034, 0037, 0038, 0040 (AIDS, Hoffner). All *Mavs* strains were IS*902* positive. One pathogenic *M.avium* strain 0033 (AIDS, Hoffner) also contained GS identifier sequence. GS identifier sequence is absent from other mycobacteria including other *M.avium*, *M.malmoense*, *M.szulgai*, *M.gordonae*, *M.chelonei*, *M.fortuitum*, *M.phlei*, as well as *E.coli*, *S.areus*, *Nocardia* sp, *Streptococcus* sp. *Shigella* sp. *Pseudomonas* sp.

Example 2:

To obtain the full sequence of GS in *Mavs* and *Mptb* we generated a genomic library of *Mavs* using the restriction endonuclease EcoRI and cloning into the vector pUC18. This achieved a representative library which was screened with ^{32}P-labelled identifier sequence yielding a positive clone containing a 17kbp insert. We constructed a restriction map of this insert and identified GS as fragments unique to *Mavs* and *Mptb* and not occurring in laboratory strains of *M.avium*. These fragments were sub-cloned into pUC18 and pGEM4Z. We identified GS contained within an 8kb region. The full nucleotide sequence was determined for GS on both DNA strands using primer walking and automated DNA sequencing. DNA sequence for GS in *Mptb* was obtained using overlapping PCR products generated using PwoDNA polymerase, a proofreading thermostable enzyme. The final DNA sequences were derived using the University of Wisconsin GCG gel assembly software package.

Example 3:

The DNA sequence of GS in *Mavs* and *Mptb* was found to be more than 99% homologous. The ORFs encoded in GS were identified using GeneRunner and DNAStar computer programmes. Eight ORFs were identified and designated GSA, GSB, GSC, GSD, GSE, GSF, GSG

and GSH. Database comparisons were carried out against the GenEMBL Database release version 48.0 (9/96), using the BLAST and BLIXEM programmes. GSA and GSB encoded proteins of 13.5kDa and 30.7kDa respectively, both of unknown functions. GSC encoded a protein of 38.4kDa with a 65% homology to the amino acid sequence of *rfbD* of *V.cholerae*, a 62% amino acid sequence homology to *gmd* of *E.coli* and a 58% homology to *gca* of *Ps.aeruginosa* which are all GDP-D-mannose dehydratases. Equivalent gene products in *H.influenzae*, *S.dysenteriae*, *Y.enterocolitica*, *N.gonorrhoea*, *K.pneumoniae* and *rfbD* in *Salmonella enterica* are all involved in 'O'-antigen processing known to be linked to pathogenicity. GSD encoded a protein of 37.1kDa which showed 58% homology at the DNA level to *wcaG* from *E.coli*, a gene involved in the synthesis and regulation of capsular polysaccharides, also related to pathogenicity. GSE was found to have a > 30% amino acid homology to *rfbT* of *V.cholerae*, involved in the transport of specific LPS components across the cell membrane. In *V.cholerae* the gene product causes a seroconversion from the Inaba to the Ogawa 'epidemic' strain. GSF encoded a protein of 30.2kDa which was homologous in the range 25-40% at the amino acid level to several glucosyl transferases such as *rfpA* of *K.pneumoniae*, *rfbB* of *K.pneumoniae*, *lgtD* of *H.influenzae*, *lsi* of *N.gonorrhoae*. In *E.coli* an equivalent gene *galE* adds β-1-3 N-acetylglucosamine to galactose, the latter only found in 'O' and 'M' antigens which are also related to pathogenicity. GSH comprising the ORFs GSH_1 and GSH_2 encodes a protein totalling about 60kDa which is a putative transposase with a 40 - 43% homology at the amino acid level to the equivalent gene product of IS*21* in *E.coli*. This family of insertion sequences is broadly distributed amongst gram negative bacteria and is responsible for mobility and transposition of genetic elements. An IS*21*- like element in *B.fragilis* is split either side of the β-lactamase gene controlling its activation and expression. We programmed an *E.coli* S30 cell-free extract with plasmid DNA containing the ORF GSH under the control of a *lac* promoter in the presence of a ^{35}S-methionine, and demonstrated the translation of an abundant 60kDa protein. The proteins homologous to GS encoded in other organisms are in general highly antigenic. Thus the proteins encoded by the ORFs

in GS may be used in immunoassays of antibody or cell mediated immuno-reactivity for diagnosing infections caused by mycobacteria, particularly *Mptb, Mavs and Mtb*. Enhancement of host immune recognition of GS encoded proteins by vaccination using naked specific DNA or recombinant GS proteins, may be used in the prevention and treatment of infections caused by *Mptb*, *Mavs* and *Mtb* in humans and animals. Mutation or deletion of all or some of the ORFs A to H in GS may be used to generate attenuated strains of *Mptb, Mavs* or *Mtb* with lower pathogenicity for use as living or killed vaccines in humans and animals. Such vaccines are particularly relevant to Johne's disease in animals, to diseases caused by *Mptb* in humans such as Crohn's disease, and to the management of tuberculosis especially where the disease is caused by multiple drug-resistant organisms.

SEQUENCE LISTING

Seq. ID No.1

```
5'- 1 GATCCAACTA AACCCGATGG AACCCCGCGC AAACTATTGG ACGTCTCCGC GCTACGCAGT
   61 TGGGTTGGCG CCCGCGAATC GCACTGAAAG AGGGCATCGA TGCAACGGTG TCGTGGTACC
  121 GCACAAATGC CGATGCCGTG AGGAGGTAAA GCTGCGGGCC GGCCGATGTT ATCCCTCCGG
  181 CCGGACGGGT AGGGCGACCT GCCATCGAGT GGTACGGCAG TCGCCTGGCC GGCGAGGCGC
  241 ATGGCCTATG TGAGTATCCC ATAGCCTGGC TTGGCTCGCC CCTACGCATT ATCAGTTGAC
  301 CGCTTTCGCG CCACGTCGCA GGCTTGCGGC AGCATCCCGT TCAGGTCTCC TCATGGTCCG
  361 GTGTGGCACG ACCACGCAAG CTCGAACCGA CTCGTTTCCC AATTTCGCAT GCTAATATCG
  421 CTCGATGGAT TTTTTGCGCA ACGCCGGCTT GATGGCTCGT AACGTTAGCA CCGAGATGCT
  481 GCGCCACTCC GAACGAAAGC GCCTATTAGT AAACCAAGTC GAAGCATACG GAGTCAACGT
  541 TGTTATTGAT GTCGGTGCTA ACTCCGGCCA GTTCGGTAGC GCTTTGCGTC GTGCAGGATT
  601 CAAGAGCCGT ATCGTTTCCT TTGAACCTCT TTCGGGGCCA TTTGCGCAAC TAACGCGCAA
  661 GTCGGCATCG GATC -3'
```

Seq. ID No.2

```
5'- 1 GATCCGATGC CGACTTGCGC GTTAGTTGCG CAAATGGCCC CGAAAGAGGT TCAAAGGAAA
   61 CGATACGGCT CTTGAATCCT GCACGACGCA AAGCGCTACC GAACTGGCCG GAGTTAGCAC
  121 CGACATCAAT AACAACGTTG ACTCCGTATG CTTCGACTTG GTTTACTAAT AGGCGCTTTC
  181 GTTCGGAGTG GCGCAGCATC TCGGTGCTAA CGTTACGAGC CATCAAGCCG GCGTTGCGCA
  241 AAAAATCCAT CGAGCGATAT TAGCATGCGA AATTGGGAAA CGAGTCGGTT CGAGCTTGCG
  301 TGGTCGTGCC ACACCGGACC ATGAGGAGAC CTGAACGGGA TGCTGCCGCA AGCCTGCGAC
  361 GTGGCGCGAA AGCGGTCAAC TGATAATGCG TAGGGGCGAG CCAAGCCAGG CTATGGGATA
  421 CTCACATAGG CCATGCGCCT CGCCGGCCAG GCGACTGCCG TACCACTCGA TGGCAGGTCG
  481 CCCTACCCGT CCGGCCGGAG GGATAACATC GGCCGGCCCG CAGCTTTACC TCCTCACGGC
  541 ATCGGCATTT GTGCGGTACC ACGACACCGT TGCATCGATG CCCTCTTTCA GTGCGATTCG
  601 CGGGCGCCAA CCCAACTGCG TAGCGCGGAG ACGTCCAATA GTTTGCGCGG GGTTCCATCG
  661 GGTTTAGTTG GATC -3'
```

Seq. ID No.3

```
   1  GAATTCTGGG TTGGAGACGA CGTCGAACTC CTGGTCGGTC TTGCTTCGAA
  51  TGATCGCTGT GATCTGGTCG GCGGTGCCGA CAGGAACCGT CGACTTGTCG
 101  ACGATCACCT TGTACCGGTC GATGTATGAC CCAATGTCGT CCGCAACCGA
 151  GAAGACGTAC GTCAGGTCCG CCGCCCCGCT TTCACCCATG GGCGTCGGGA
 201  CGGCGATGAA AATGACGTCC GCGTGCTCGA TTCCGCGTTG CCGGTCGGTG
 251  GTGAAGTCAA TCAGCCCGTT CTCACGGTTC CTCGCAATCA ACTCCCAACC
 301  CGGGCTCGAA AATCGGGACA CTGCCTGCGA GGAGCAAATC GATCTTGGCC
 351  TGATCGATAT CGACACAGAC GACATCGTTG CCGCTATCCG CGAGACAGGC
 401  GCCCGTGACG AGGCCTACAT AGCCTGATCC GACCACCGAA ATTTTCAAGA
 451  TGACCCCTTC AAGTCCCCGA TCGGTCGACG ACCATACTGC CGCAACTCTG
 501  TACCCTCCGT GGGTAATTCG CATGTCGCGT TCGTAAGGAG CAGCCAGCGA
 551  GTCGGGGACG TTCGGTGAGA GAGTCGCAGG ACTACGAGGT TGCCGGTGCG
 601  ATACATCACA GTGTTGCGTC TGTCGGCAAC GATGCAGCAA GAACCCACGG
 651  GGCAGCCCTG AACTGCGCGC ATGACCGGTC CTTGTCCTGG CACCTTTGAT
 701  CGGCCACCGC TTCCATGCGA ACATGACCGG AATCCATAGC GCGTGGTCAA
 751  GCAGCGGGGA GGTAGACGTC GGTGTCATCT GCTCCAACCG TGTCGGTGAT
 801  AACGATTTCG CTGAACGATC TCGAGGGATT GAAAAGCACC GTGGAGAGCG
 851  TTCGCGCGCA GCGCTATGGG GGGCGAATCG AGCACATCGT CATCGACGGT
 901  GGATCGGGCG ACGCCGTCGT GGAGTATCTG TCCGGCGATC CTGGCTTTGC
 951  ATATTGGCAA TCTCAGCCCG ACAACGGGAG ATATGACGCG ATGAATCAGG
1001  GCATTGCCCA TTCGTCGGGC GACCTGTTGT GGTTTATGCA CTCCACGGAT
1051  CGTTTCTCCG ATCCAGATGC AGTCGCTTCC GTGGTGGAGG CGCTCTCGGG
1101  GCATGGACCA GTACGTGATT TGTGGGGTTA CGGGAAAAAC AACCTTGTCG
1151  GACTCGACGG CAAACCACTT TTCCCTCGGC CGTACGGCTA TATGCCGTTT
1201  AAGATGCGGA AATTTCTGCT CGGCGCGACG GTTGCGCATC AGGCGACATT
1251  CTTCGGCGCG TCGCTGGTAG CCAAGTTGGG CGGTTACGAT CTTGATTTTG
1301  GACTCGAGGC GGACCAGCTG TTCATCTACC GTGCCGCACT AATACGGCCT
1351  CCCGTCACGA TCGACCGCGT GGTTTGCGAC TTCGATGTCA CGGGACCTGG
1401  TTCAACCCAG CCCATCCGTG AGCACTATCG GACCCTGCGG CGGCTCTGGG
1451  ACCTGCATGG CGACTACCCG CTGGGTGGGC GCAGAGTGTC GTGGGCTTAC
1501  TTGCGTGTGA AGGAGTACTT GATTCGGGCC GACCTGGCCG CATTCAACGC
1551  GGTAAAGTTC TTGCGAGCGA AGTTCGCCAG AGCTTCGCGG AAGCAAAATT
1601  CATAGAAACC AACTTCTACT GCCTGACCTG AGCAGCGCCG AGGCGCGCAG
1651  CGCGATCAGT GCGACCTGAA CGGCCAGGTG GAAAGCGCCA CCGATCCCGG
1701  CACCGAGTGC CTGACGCTTC GGATCCCTTG CACCACAACG AGAGTGAGAG
1751  CGCCATGATG AGGAAATATC GGCTGGGCGG AGTCAACGCC GGAGTGACAA
1801  AAGTGAGAAC CCGGTGAAGC GAGCGCTTAT AACAGGGATC ACGGGGCAGG
1851  ATGGTTCCTA CCTCGCCGAG CTACTACTGA GCAAGGGATA CGAGGTTCAC
1901  GGGCTCGTTC GTCGAGCTTC GACGTTTAAC ACGTCGCGGA TCGATCACCT
1951  CTACGTTGAC CCACACCAAC CGGGCGCGCG CTTGTTCTTG CACTATGCAG
2001  ACCTCACTGA CGGCACCCGG TTGGTGACCC TGCTCAGCAG TATCGACCCG
2051  GATGAGGTCT ACAACCTCGC AGCGCAGTCC CATGTGCGCG TCAGCTTTGA
2101  CGAGCCAGTG CATACCGGAG ACACCACCGG CATGGGATCG ATCCGACTTC
2151  TGGAAGCAGT CCGCCTTTCT CGGGTGGACT GCCGGTTCTA TCAGGCTTCC
2201  TCGTCGGAGA TGTTCGGCGC ATCTCCGCCA CCGCAGAACG AATCGACGCC
2251  GTTCTATCCC CGTTCGCCAT ACGGCGCGGC CAAGGTCTTC TCGTACTGGA
2301  CGACTCGCAA CTATCGAGAG GCGTACGGAT TATTCGCAGT GAATGGCATC
2351  TTGTTCAACC ATGAGTCCCC CCGGCGCGGC GAGACTTTCG TGACCCGAAA
2401  GATCACGCGT GCCGTGGCGC GCATCCGAGC TGGCGTCCAA TCGGAGGTCT
2451  ATATGGGCAA CCTCGATGCG ATCCGCGACT GGGGCTACGC GCCCGAATAT
2501  GTCGAGGGGA TGTGGAGGAT GTTGCAAGCG CCTGAACCTG ATGACTACGT
```

```
2551  CCTGGCGACA GGGCGTGGTT ACACCGTACG TGAGTTCGCT CAAGCTGCTT
2601  TTGACCATGT CGGGCTCGAC TGGCAAAAGC GCGTCAAGTT TGACGACCGC
2651  TATTTGCGTC CCACCGAGGT CGATTCGCTA GTAGGAGATG CCGACAAGGC
2701  GGCCCAGTCA CTCGGCTGGA AAGCTTCGGT TCATACTGGT GAACTCGCGC
2751  GCATCATGGT GGACGCGGAC ATCGCCGCGT TGGAGTGCGA TGGCACACCA
2801  TGGATCGACA CGCCGATGTT GCCTGGTTGG GGCAGAGTAA GTTGACGACT
2851  ACACCTGGGC CTCTGGACCG CGCAACGCCC GTGTATATCG CCGGTCATCG
2901  GGGGCTGGTC GGCTCAGCGC TCGTACGTAG ATTTGAGGCC GAGGGGTTCA
2951  CCAATCTCAT TGTGCGATCA CGCGATGAGA TTGATCTGAC GGACCGAGCC
3001  GCAACGTTTG ATTTTGTGTC TGAGACAAGA CCACAGGTGA TCATCGATGC
3051  GGCCGCACGG GTCGGCGGCA TCATGGCGAA TAACACCTAT CCCGCGGACT
3101  TCTTGTCCGA AAACCTCCGA ATCCAGACCA ATTTGCTCGA CGCAGCTGTC
3151  GCCGTGCGTG TGCCGCGGCT CCTTTTCCTC GGTTCGTCAT GCATCTACCC
3201  GAAGTACGCT CCGCAACCTA TCCACGAGAG TGCTTTATTG ACTGGCCCTT
3251  TGGAGCCCAC CAACGACGCG TATGCGATCG CCAAGATCGC CGGTATCCTG
3301  CAAGTTCAGG CGGTTAGGCG CCAATATGGG CTGGCGTGGA TCTCTGCGAT
3351  GCCGACTAAC CTCTACGGAC CCGGCGACAA CTTCTCCCCG TCCGGGTCGC
3401  ATCTCTTGCC GGCGCTCATC CGTCGATATG AGGAAGCCAA AGCTGGTGGT
3451  GCAGAAGAGG TGACGAATTG GGGGACCGGT ACTCCGCGGC GCGAACTTCT
3501  GCATGTCGAC GATCTGGCGA GCGCATGCCT GTTCCTTTTG GAACATTTCG
3551  ATGGTCCGAA CCACGTCAAC GTGGGCACCG GCGTCGATCA CAGCATTAGC
3601  GAGATCGCAG ACATGGTCGC TACAGCGGTG GGCTACATCG GCGAAACACG
3651  TTGGGATCCA ACTAAACCCG ATGGAACCCC GCGCAAACTA TTGGACGTCT
3701  CCGCGCTACG CGAGTTGGGT TGGCGCCCGC GAATCGCACT GAAAGACGGC
3751  ATCGATGCAA CGGTGTCGTG GTACCGCACA AATGCCGATG CCGTGAGGAG
3801  GTAAAGCTGC GGGTCGGCCG ATGTTATCCC TCCGGCCGGA CGGGTGGGGC
3851  GACCTGCCGT CGAGTGGTAC GGCAGTCGCC TGGCCGGCGA GGCGCGTGGC
3901  CTATGGGAGT ATCCAATAGC CTGGCTTGGC TCGCCCCTAC GCATTATCAG
3951  TTGACCGCTT TCGCGCCAGC TCGCAGGCTT GCGGCAGCAT CCCGTTCAGG
4001  TCTCCTCATG GTCCGGTGTG GCACGACCAC GCAAGCTCGA ACCGACTCGT
4051  TTCCCAATTT CGCATGCTAA TATCGCTCGA TGGATTTTTT GCGCAACGCC
4101  GGCTTGATGG CTCGTAACGT TAGTACCGAG ATGCTGCGCC ACTTCGAACG
4151  AAAGCGCCTA TTAGTAAACC AATTCAAAGC ATACGGAGTC AACGTTGTTA
4201  TTGATGTCGG TGCTAACTCC GGCCAGTTCG GTAGCGCTTT GCGTCGTGCA
4251  GGATTCAAGA GCCGTATCGT TTCCTTTGAA CCTCTTTCGG GGCCATTTGC
4301  GCAACTAACG CGCAAGTCGG CATCGGATCC ACTATGGGAG TGTCACCAGT
4351  ATGCCCTAGG CGACGCCGAT GAGACGATTA CCATCAATGT GGCAGGCAAT
4401  GCGGGGGCAA GTAGTTCCGT GCTGCCGATG CTTAAAAGTC ATCAAGATGC
4451  CTTTCCTCCC GCGAATTATA TTGGCACCGA AGACGTTGCA ATACACCGCC
4501  TTGATTCGGT TGCATCAGAA TTTCTGAACC CTACCGATGT TACTTTCCTG
4551  AAGATCGACG TACAGGGTTT CGAGAAGCAG GTTATCACGG GCAGTAAGTC
4601  AACGCTTAAC GAAAGCTGCG TCGGCATGCA ACTCGAACTT TCTTTTATTC
4651  CGTTGTACGA AGGTGACATG CTGATTCATG AAGCGCTTGA ACTTGTCTAT
4701  TCCCTAGGTT TCAGACTGAC GGGTTTGTTG CCCGGCTTTA CGGATCCGCG
4751  CAATGGTCGA ATGCTTCAAG CTGACGGCAT TTTCTTCCGT GGGGACGATT
4801  GACATAAATG CTCCGTCGGC ACCCTGCCGG TATCCAAACG GGCGATCTGG
4851  TGAGCCGGCC TCCCGGGCAC CTAATCGACT ATCTAAATTG AGGCGGCCGC
4901  GACGTGCGGC ACGAACAGGT GGCCGGCTGC TAGCGTTACA CACGTCATGA
4951  CTGCGCCAGT GTTCTCGATA ATTATCCCTA CCTTCAATGC AGCGGTGACG
5001  CTGCAAGCCT GCCTCGGAAG CATCGTCGGG CAGACCTACC GGGAAGTGGA
5051  AGTGGTCCTT GTCGACGGCG GTTCGACCGA TCGGACCCTC GACATCGCGA
5101  ACAGTTTCCG CCCGGAACTC GGCTCGCGAC TGGTCGTTCA CAGCGGGCCC
5151  GATGATGGCC CCTACGACGC CATGAACCGC GGCGTCGGCG TGGCCACAGG
```

```
5201  CGAATGGGTA CTTTTTTTAG GCGCCGACGA CACCCTCTAC GAACCAACCA
5251  CGTTGGCCCA GGTAGCCGCT TTTCTCGGCG ACCATGCGGC AAGCCATCTT
5301  GTCTATGGCG ATGTTGTGAT GCGTTCGACG AAAAGCCGGC ATGCCGGACC
5351  TTTCGACCTC GACCGCCTCC TATTTGAGAC GAATTTGTGC CACCAATCGA
5401  TCTTTTACCG CCGTGAGCTT TTCGACGGCA TCGGCCCTTA CAACCTGCGC
5451  TACCGAGTCT GGGCGGACTG GGACTTCAAT ATTCGCTGCT TCTCCAACCC
5501  GGCGCTGATT ACCCGCTACA TGGACGTCGT GATTTCCGAA TACAACGACA
5551  TGACCGGCTT CAGCATGAGG CAGGGGACTG ATAAAGAGTT CAGAAAACGG
5601  CTGCCAATGT ACTTCTGGGT TGCAGGGTGG GAGACTTGCA GGCGCATGCT
5651  GGCGTTTTTG AAAGACAAGG AGAATCGCCG TCTGGCCTTG CGTACGCGGT
5701  TGATAAGGGT TAAGGCCGTC TCCAAAGAAC GAAGCGCAGA ACCGTAGTCG
5751  CGGATCCACA TTGGACTTCT TTAACGCGTT TGCGTCCTGA TCCACCTTTC
5801  AAGCCCGTTC CGCGTAACGC GGCGCGCAGA GAGTGGTCGC ATATCGCATC
5851  ACTGTTCTCG TGCCAGTGCT TGGAAAGCGT CGAGCACTCT GGTTCGCGTT
5901  CTTGACGTTC GCGCCCGCTC CTAGAGGTAG CGTGTCACGT GACTGAAGCC
5951  AATGAGTGCA ACTCGGCGTC GCGAAAGGTT TCAGTCGCGG TTGAGCAAGA
6001  CACCGCAAGA CTACTGGAGT GCGTGCACAA GCGCCTCCAG CTCGCGGCTG
6051  AAAGCGGATG CAAAGGGATT CGAAGCTTGA GCAACATGCG AAGGGGAGAA
6101  CGGCCTATGA GGCTGGGACA GGTTTTCGAT CCGCGCGCGA ATGCACTGTC
6151  AATGGCCAAG TAGAAGTCCC CGCTGGTGGC CAGCAGAAGT CCCCACTCCG
6201  CTGCGGGTGG TTGGCTAATT CTTGGCGGCT CCCTTCTTGT GGTCGGCGTG
6251  GCGCATCCGG TAGGACTCGC CGGAGGTGAC GACGATGCTG GCGTGGTGCA
6301  GCAGCCGATC GAGGATGCTG GCGGCGGTGG TGTGCTCGGG CAGGAATCGC
6351  CCCCATTGTT CGAAGGGCCA ATGCGAGGCG ATGGCCAGGG AGCGGCGCTC
6401  GTAGCCGGCA GCCACGAGCC GGAACAACAG TTGAGTCCCG GTGTCGTCGA
6451  GCGGGGCGAA GCCGATCTCG TCCAAGATGA CCAGATCCGC GCGGAGCAGG
6501  GTGTCGATGA TCTTGCCGAC GGTGTTGTCG GCCAGGCCGC GGTAGAGGAC
6551  CTCGATCAGG TCGGCGGCGG TGAAGTAGCG GACTTTGAAT CCGGCGTGGA
6601  CGGCAGCGTG CCCGCAGCCG ATGAGCAGGT GACTTTTGCC CGTACCAGGT
6651  GGGCCAATGA CCGCCAGGTT CTGTTGTGCC CGAATCCATT CCAGGCTCGA
6701  CAGGTAGTCG AACGTGGCTG CGGTGATCGA CGATCCGGTG ACGTCGAACC
6751  CGTCGAGGGT CTTGGTGACC GGGAAGGCTG CGGCCTTGAG ACGGTTGGCG
6801  GTGTTGGAGG CATCGCGGGC AGCGATCTCG GCCTCAACCA ACGTCCGCAG
6851  GATCTCCTCC GGTGTCCAGC GTTGCGTCTT GGCGACTTGC AACACCTCGG
6901  CGGCGTTGCG GCGCACCGTG GCCAGCTTCA ACCGCCGCAG CGCCGCGTCA
6951  AGGTCAGCAG CCAGCGGTGC CGCCGAGGAC GGTGCCACCG GCTTGGCAGC
7001  GGTGGTCATG AGGCCGTCCC GTCGGTGGTG TTGATCTTGT AGGCCTCCAA
7051  CGAGCGGGTC TCGACGGTGG GCAGATCGAG CACGAGTGCG TCGCCGGCGG
7101  GGCGGGGTTG TGGGGTGCCG GCGCCGGCGG CCAGGATCGA GCGCACGTCG
7151  GCAGCGCGGA ACCGGCGAAA CGCAACCGCC CGGCGCAGCG CGTCAATCAA
7201  AGCCTGTTCG CCGTGGGCGG CGCCAAGGCC GAGCAGAATG TCGAGTTCGG
7251  ATTTCAGTCG GGTGTTGCCG ATCGCAGCAG CACCGACGAG GAACTGCTGC
7301  GCTTCGGTTC CCAATGCGCA GAATCGTTTC TCTGCTTGGG TTTTCGGGCG
7351  AGGACCACGC GAGGGTGCGG GTCTGGGTCC GTCGTAGTGT TCATCGAGGA
7401  TGGACACCTC ACCTGGGCTG ACGAGCTCGT GCTCGGCCAC GATCACACCG
7451  GTCGCAGGTT CCAACAGGAT CAGGGCGCCA TGATCGACCA CCACCGCCAC
7501  GGTGGCACCG ACGAGCCGCT GAGGCACCGA GTAACGAGCT GAGCCGTAAC
7551  GGATGCACGA GAGGCCGTCG ACCTTACGGC GCACCGACCC CGAGCCGATC
7601  GTCGGCCGCA GCGAGGGCAG CTCCCTCAAG ACGGTGCGCT CGTCAACCAA
7651  GCGATCGTTG GGCACGGCGC AGATCTCCGA GTGGACCGTG GCATTGACCT
7701  CGGCGCACCA TAGTTGCGCC TGGGCGTTGA GGGCACGTAG GTCGACCTGC
7751  TCACCGGCTA ACGCAGCTTC GGTCAGCAGC GGCACCGCAA GGTCGTCCTG
7801  AGCGTAGCCA CAGAGGTTCT CCACGATGCC CTTCGATTGC GGATCCGCAC
```

7851 CGTGGCAGAA GTCCGGAACG AAGCCATAGT GGGACGCGAA TCGCACATAA
7901 TCCGGTGTTG GAACAACAAC ATTGGCGACG ACACCACCTT TGAGGCAGCC
7951 CATCCGGTCG GCCAGGATCT TGGCCGGAAC CCCACCGATC GCCTC

Seq. ID No.4

1 TTCTACTGCC TGACCTGAGC AGCGCCGAGG CGCGCAGCGC GATCACTGCG ACCTGAATGG
61 CCAGGTGGAA AGCGCCACCG ATCCCGGCAC CGAGTGCCTG ACGATTCGGA TCCCTTGCAC
121 CACAACGAGA GTGAGACCGC CATGATGACG AAATATCGGC TGGGCGGAGT CAACGCCGGA
181 GTGACAAAAG TGAGAACCCG GTGAAGCGAG CGCTTATAAC AGGGATCACG GGGCAGGATG
241 GTTCCTACCT CGCCGAGCTA CTACTGAGCA AGGGATACGA GGTTCACGGG CTCGTTCGTC
301 GAGCTTCGAC GTTTAACACG TCGCGGATCG ATCACCTCTA CGTTGACCCA CACCAACCGG
361 GCGCGCGCTT GTTCTTGCAC TATGCAGACC TCACTGACGG CACCCGGTTG GTGACCCTGC
421 TCAGCAGTAT CGACCCGGAT GAGGTCTACA ACCTCGCAGC GCAGTCCCAT GTGCGCGTCA
481 GCTTTGACGA GCCAGTGCAT ACCGGAGACA CCACCGGCAT GGGATCGATC CGACTTCTGG
541 AAGCAGTCCG CCTTTCTCGG GTGGACTGCC GGTTCTATCA GGCTTCCTCG TCGGAGATGT
601 TCGGCGCATC TCCGCCACCG CAGAACGAAT CGACGCCGTT CTATCCCCGT TCGCCATACG
661 GCGCGGCCAA GGTCTTCTCG TACTGGACGA CTCGCAACTA TCGAGAGGCG TACGGATTAT
721 TCGCAGTGAA TGGCATCTTG TTCAACCATG AGTCCCCCCG GCGCGGCGAG ACTTTCGTGA
781 CCCGAAAGAT CACGCGTGCC GTGGCGCGCA TCCGAGCTGG CGTCCAATCG GAGGTCTATA
841 TGGGCAACCT CGATGCGATC CGCGACTGGG GCTACGCGCC CGAATATGTC GAGGGGATGT
901 GGAGGATGTT GCAAGCGCCT GAACCTGATG ACTACGTCCT GGCGACAGGG CGTGGTTACA
961 CCGTACGTGA GTTCGCTCAA GCTGCTTTTG ACCACGTCGG GCTCGACTGG CAAAAGCACG
1021 TCAAGTTTGA CGACCGCTAT TTGCGCCCCA CCGAGGTCGA TTCGCTAGTA GGAGATGCCG
1081 ACAGGGCGGC CCAGTCACTC GGCTGGAAAG CTTCGGTTCA TACTGGTGAA CTCGCGCGCA
1141 TCATGGTGGA CGCGGACATC GCCGCGTCGG AGTGCGATGG CACACCATGG ATCGACACGC
1201 CGATGTTGCC TGGTTGGGGC GGAGTAAGTT GACGACTACA CCTGGGCCTC TGGACCGCGC
1261 AACGCCCGTG TATATCGCCG GTCATCGGGG GCTGGTCGGC TCAGCGCTCG TACGTAGATT
1321 TGAGGCCGAG GGGTTCACCA ATCTCATTGT GCGATCACGC GATGAGATTG ATCTGACGGA
1381 CCGAGCCGCA ACGTTTGATT TTGTGTCTGA GACAAGACCA CAGGTGATCA TCGATGCGGC
1441 CGCACGGGTC GGCGGCATCA TGGCGAATAA CACCTATCCC GCGGACTTCT TGTCCGAAAA
1501 CCTCCGAATC CAGACCAATT TGCTCGACGC AGCTGTCGCC GTGCGTGTGC CGCGGCTCCT
1561 TTTCCTCGGT TCGTCATGCA TCTACCCGAA GTACGCTCCG CAACCTATCC ACGAGAGTGC
1621 TTTATTGACT GGCCCTTTGG AGCCCACCAA CGACGCGTAT GCGATCGCCA AGATCGCCGG
1681 TATCCTGCAA GTTCAGGCGG TTAGGCGCCA ATATGGGCTG GCGTGGATCT CTGCGATGCC
1741 GACTAACCTC TACGGACCCG GCGACAACTT CTCCCCGTCC GGGTCGCATC TCTTGCCGGC
1801 GCTCATCCGT CGATATGAGG AAGCCAAAGC TGGTGGTGCA GAAGAGGTGA CGAATTGGGG
1861 GACCGGTACT CCGCGGCGCG AACTTCTGCA TGTCGACGAT CTGGCGAGCG CATGCCTGTT
1921 CCTTTTGGAA CATTTCGATG GTCCGAACCA CGTCAACGTG GGCACCGGCG TCGATCACAG
1981 CATTAGCGAG ATCGCAGACA TGGTCGCTAC GGCGGTGGGC TACATCGGCG AAACACGTTG
2041 GGATCCAACT AAACCCGATG GAACCCCGCG CAAACTATTG GACGTCTCCG CGCTACGCGA
2101 GTTGGGTTGG CGCCCGCGAA TCGCACTGAA AGACGGCATC GATGCAACGG TGTCGTGGTA
2161 CCGCACAAAT GCCGATGCCG TGAGGAGGTA AAGCTGCGGG CCGGCCGATG TTATCCCTCC
2221 GGCCGGACGG GTAGGGCGAC CTGCCATCGA GTGGTACGGC AGTCGCCTGG CCGGCGAGGC
2281 GCATGGCCTA TGGGAGTATC CCATAGCCTG GCTTGGCTCG CCCCTACGCA TTATCAGTTG
2341 ACCGCTTTCG CGCCAGCTCG CAGGCTCGCG GCAGCATCCC GTTCAGGTCT CCTCATGGTC
2401 CGGTGTGGCA CGACCACGCA AGCTCGAACC GACTCGTTTC CCAATTTCGC ATGCTAATAT
2461 CGCTCGATGG ATTTTTTGCG CAACGCCGGC TTGATGGCTC GTAACGTTAG CACCGAGATG
2521 CTGCGCCACT TCGAACGAAA GCGCCTATTA GTAAACCAAT TCAAAGCATA CGGAGTCAAC
2581 GTTGTTATTG ATGTCGGTGC TAACTCCGGC CAGTTCGGTA GCGCTTTGCG TCGTGCAGGA
2641 TTCAAGAGCC GTATCGTTTC CTTTGAACCT CTTTCGGGGC CATTTGCGCA ACTAACGCGC
2701 GAGTCGGCAT CGGATCCACT ATGGGAGTGT CACCAGTATG CCCTAGGCGA CGCCGATGAG

```
2761 ACGATTACCA TCAATGTGGC AGGCAATGCG GGGGCAAGTA GTTCCGTGCT GCCGATGCTT
2821 AAAAGTCATC AAGATGCCTT TCCTCCCGCG AATTATATTG GCACCGAAGA CGTTGCAATA
2881 CACCGCCTTG ATTCGGTTGC ATCAGAATTT CTGAACCCTA CCGATGTTAC TTTCCTGAAG
2941 ATCGACGTAC AGGGTTTCGA GAAGCAGGTT ATCGCGGGCA GTAAGTCAAC GCTTAACGAA
3001 AGCTGCGTCG GCATGCAACT CGAACTTTCT TTTATTCCGT TGTACGAAGG TGACATGCTG
3061 ATTCATGAAG CGCTTGAACT TGTCTATTCC CTAGGTTTCA GACTGACGGG TTTGTTGCCC
3121 GGATTTACGG ATCCGCGCAA TGGTCGAATG CTTCAAGCTG ACGGCATTTT CTTCCGTGGG
3181 GACGATTGAC ATAAATGCTT GCGTCGGCAC CCTGCCGGTA TCCAAACGGG CGATCTGGTG
3241 AGCCGGCCTC CCGGGCACCT AATCGACTAT CTAAATTGAG GCGGCCGCGA CGTGCGGCAC
3301 GAACAGGTGG CCGGCTGCTA GCGTTACACA CGTCATGACT GCGCCAGTGT TCTCGATAAT
3361 TATCCCTACC TTCAATGCAG CGGTGACGCT GCAAGCCTGC CTCGGAAGCA TCGTCGGGCA
3421 GACCTACCGG GAAGTGGAAG TGGTCCTTGT CGACGGCGGT TCGACCGATC GGACCCTCGA
3481 CATCGCGAAC AGTTTCCGCC CGGAACTCGG CTCGCGACTG GTCGTTCACA GCGGGCCCGA
3541 TGATGGCCCC TACGACGCCA TGAACCGCGG CGTCGGCGTA GCCACAGGCG AATGGGTACT
3601 TTTTTTAGGC GCCGACGACA CCCTCTACGA ACCAACCACG TTGGCCCAGG TAGCCGCTTT
3661 TCTCGGCGAC CATGCGGCAA GCCATCTTGT CTATGGCGAT GTTGTGATGC GTTCGACGAA
3721 AAGCCGGCAT GCCGGACCTT TCGACCTCGA CCGCCTCCTA TTTGAGACGA ATTTGTGCCA
3781 CCAATCGATC TTTTACCGCC GTGAGCTTTT CGACGGCATC GGCCCTTACA ACCTGCGCTA
3841 CCGAGTCTGG GCGGACTGGG ACTTCAATAT TCGCTGCTTC TCCAACCCGG CGCTGATTAC
3901 CCGCTACATG GACGTCGTGA TTTCCGAATA CAACGACATG ACCGGCTTCA GCATGAGGCA
3961 GGGGACTGAT AAAGAGTTCA GAAAACGGCT GCCAATGTAC TTCTGGGTTG CAGGGTGGGA
4021 GACTTGCAGG CGCATGCTGG CGTTTTTGAA AGACAAGGAG AATCGCCGTC TGGCCTTGCG
4081 TACGCGGTTG ATAAGGGTTA AGGCCGTCTC CAAAGAACGA AGCGCAGAAC CGTAGTCGCG
4141 GATCCACATT GGACTTCTTT AACGCGTTTG CGTCCTGATC CACCTTTCAA CCCCGTTCCG
4201 CGTGACGCGG CGCGCAGAGA GTGGTCGCAT ATCGCGTCAC TGTTCTCGTG CCAGTGCTTG
4261 GAAAGCGTCG AGCACTCTGG TTCGCGTTCT TGACGTTCGC GCCCGCCCCT AGAGGTAGCG
4321 TGTCACGTGA CTGAAGCCAA TGAGTGCAAC TCGGCGTCGC GAAAGGTTTC AGTCGCGGTT
4381 GAGCAAGACA CCGCAAGACT ACTGGAGTGC GTGCACAAGC GCCTCCAGCT CACGG
```

Seq. ID No.5

```
   1 atgatcgctg tgatctggtc ggcggtgccg acaggaaccg tcgacttgtc gacgatcacc
  61 ttgtaccggt cgatgtatga cccaatgtcg tccgcaaccg agaagacgta cgtcaggtcc
 121 gccgccccgc tttcacccat gggcgtcggg acggcgatga aaatgacgtc cgcgtgctcg
 181 attccgcgtt gccggtcggt ggtgaagtca atcagcccgt tctcacggtt cctcgcaatc
 241 aactcccaac ccgggctcga aaatcgggac actgcctgcg aggagcaaat cgatcttggc
 301 ctgatcgata tcgacacaga cgacatcgtt gccgctatcc gcgagacagg cgcccgtgac
 361 gaggcctaca tagcctga
```

Seq. ID No.6

```
   1  M I A V I W S A V P T G T V D L S T I T L Y R S M Y D P M S
  31  S A T E K T Y V R S A A P L S P M G V G T A M K M T S A C S
  61  I P R C R S V V K S I S P F S R F L A I N S Q P G L E N R D
  91  T A C E E Q I D L G L I D I D T D D I V A A I R E T G A R D
 121  E A Y I A
```

Seq. ID No.7

```
   1 gtgtcatctg ctccaaccgt gtcggtgata acgatttcgc tgaacgatct cgagggattg
  61 aaaagcaccg tggagagcgt tcgcgcgcag cgctatgggg ggcgaatcga gcacatcgtc
 121 atcgacggtg gatcgggcga cgccgtcgtg gagtatctgt ccggcgatcc tggctttgca
 181 tattggcaat ctcagcccga caacgggaga tatgacgcga tgaatcaggg cattgcccat
 241 tcgtcgggcg acctgttgtg gtttatgcac tccacggatc gtttctccga tccagatgca
 301 gtcgcttccg tggtggaggc gctctcgggg catggaccag tacgtgattt gtggggttac
 361 gggaaaaaca accttgtcgg actcgacggc aaaccacttt tccctcggcc gtacggctat
 421 atgccgttta agatgcggaa atttctgctc ggcgcgacgg ttgcgcatca ggcgacattc
 481 ttcggcgcgt cgctggtagc caagttgggc ggttacgatc ttgattttgg actcgaggcg
 541 gaccagctgt tcatctaccg tgccgcacta atacggcctc ccgtcacgat cgaccgcgtg
 601 gtttgcgact tcgatgtcac gggacctggt tcaacccagc ccatccgtga gcactatcgg
 661 accctgcggc ggctctggga cctgcatggc gactacccgc tgggtgggcg cagagtgtcg
 721 tgggcttact tgcgtgtgaa ggagtacttg attcgggccg acctggccgc attcaacgcg
 781 gtaaagttct tgcgagcgaa gttcgccaga gcttcgcgga agcaaaattc atag
```

Seq. ID No.8

```
   1  V S S A P T V S V I T I S L N D L E G L K S T V E S V R A Q
  31  R Y G G R I E H I V I D G G S G D A V V E Y L S G D P G F A
  61  Y W Q S Q P D N G R Y D A M N Q G I A H S S G D L L W F M H
  91  S T D R F S D P D A V A S V V E A L S G H G P V R D L W G Y
 121  G K N N L V G L D G K P L F P R P Y G Y M P F K M R K F L L
 151  G A T V A H Q A T F F G A S L V A K L G G Y D L D F G L E A
 181  D Q L F I Y R A A L I R P P V T I D R V V C D F D V T G P G
 211  S T Q P I R E H Y R T L R R L W D L H G D Y P L G G R R V S
 241  W A Y L R V K E Y L I R A D L A A F N A V K F L R A K F A R
 271  A S R K Q N S
```

Seq. ID No.9

```
   1 gtgaagcgag cgcttataac agggatcacg gggcaggatg gttcctacct cgccgagcta
  61 ctactgagca agggatacga ggttcacggg ctcgttcgtc gagcttcgac gtttaacacg
 121 tcgcggatcg atcacctcta cgttgaccca caccaaccgg gcgcgcgctt gttcttgcac
 181 tatgcagacc tcactgacgg cacccggttg gtgaccctgc tcagcagtat cgacccggat
 241 gaggtctaca acctcgcagc gcagtcccat gtgcgcgtca gctttgacga gccagtgcat
 301 accggagaca ccaccggcat gggatcgatc cgacttctgg aagcagtccg cctttctcgg
 361 gtggactgcc ggttctatca ggcttcctcg tcggagatgt tcggcgcatc tccgccaccg
 421 cagaacgaat cgacgccgtt ctatccccgt tcgccatacg gcgcggccaa ggtcttctcg
 481 tactggacga ctcgcaacta tcgagaggcg tacggattat tcgcagtgaa tggcatcttg
 541 ttcaaccatg agtccccccg gcgcggcgag actttcgtga cccgaaagat cacgcgtgcc
 601 gtggcgcgca tccgagctgg cgtccaatcg gaggtctata tgggcaacct cgatgcgatc
 661 cgcgactggg gctacgcgcc cgaatatgtc gaggggatgt ggaggatgtt gcaagcgcct
 721 gaacctgatg actacgtcct ggcgacaggg cgtggttaca ccgtacgtga gttcgctcaa
 781 gctgcttttg accatgtcgg gctcgactgg caaaagcgcg tcaagtttga cgaccgctat
 841 ttgcgtccca ccgaggtcga ttcgctagta ggagatgccg acaaggcggc ccagtcactc
 901 ggctggaaag cttcggttca tactggtgaa ctcgcgcgca tcatggtgga cgcggacatc
 961 gccgcgttgg agtgcgatgg cacaccatgg atcgacacgc cgatgttgcc tggttggggc
1021 agagtaagtt ga
```

Seq. ID No.10

```
  1 V K R A L I T G I T G Q D G S Y L A E L L L S K G Y E V H G
 31 L V R R A S T F N T S R I D H L Y V D P H Q P G A R L F L H
 61 Y A D L T D G T R L V T L L S S I D P D E V Y N L A A Q S H
 91 V R V S F D E P V H T G D T T G M G S I R L L E A V R L S R
121 V D C R F Y Q A S S S E M F G A S P P P Q N E S T P F Y P R
151 S P Y G A A K V F S Y W T T R N Y R E A Y G L F A V N G I L
181 F N H E S P R R G E T F V T R K I T R A V A R I R A G V Q S
211 E V Y M G N L D A I R D W G Y A P E Y V E G M W R M L Q A P
241 E P D D Y V L A T G R G Y T V R E F A Q A A F D H V G L D W
271 Q K R V K F D D R Y L R P T E V D S L V G D A D K A A Q S L
301 G W K A S V H T G E L A R I M V D A D I A A L E C D G T P W
331 I D T P M L P G W G R V S
```

Seq. ID No.11

```
   1 gtgaagcgag cgcttataac agggatcacg gggcaggatg gttcctacct cgccgagcta
  61 ctactgagca agggatacga ggttcacggg ctcgttcgtc gagcttcgac gtttaacacg
 121 tcgcggatcg atcacctcta cgttgaccca caccaaccgg gcgcgcgctt gttcttgcac
 181 tatgcagacc tcactgacgg cacccggttg gtgaccctgc tcagcagtat cgacccggat
 241 gaggtctaca acctcgcagc gcagtcccat gtgcgcgtca gctttgacga gccagtgcat
 301 accggagaca ccaccggcat gggatcgatc cgacttctgg aagcagtccg cctttctcgg
 361 gtggactgcc ggttctatca ggcttcctcg tcggagatgt tcggcgcatc tccgccaccg
 421 cagaacgaat cgacgccgtt ctatccccgt tcgccatacg gcgcggccaa ggtcttctcg
 481 tactggacga ctcgcaacta tcgagaggcg tacggattat tcgcagtgaa tggcatcttg
 541 ttcaaccatg agtccccccg gcgcggcgag actttcgtga cccgaaagat cacgcgtgcc
 601 gtggcgcgca tccgagctgg cgtccaatcg gaggtctata tgggcaacct cgatgcgatc
 661 cgcgactggg gctacgcgcc cgaatatgtc gaggggatgt ggaggatgtt gcaagcgcct
 721 gaacctgatg actacgtcct ggcgacaggg cgtggttaca ccgtacgtga gttcgctcaa
 781 gctgcttttg accacgtcgg gctcgactgg caaaagcacg tcaagtttga cgaccgctat
 841 ttgcgcccca ccgaggtcga ttcgctagta ggagatgccg acagggcggc ccagtcactc
 901 ggctggaaag cttcggttca tactggtgaa ctcgcgcgca tcatggtgga cgcggacatc
 961 gccgcgtcgg agtgcgatgg cacaccatgg atcgacacgc cgatgttgcc tggttggggc
1021 ggagtaagtt ga
```

Seq. ID No.12

```
  1 V K R A L I T G I T G Q D G S Y L A E L L L S K G Y E V H G
 31 L V R R A S T F N T S R I D H L Y V D P H Q P G A R L F L H
 61 Y A D L T D G T R L V T L L S S I D P D E V Y N L A A Q S H
 91 V R V S F D E P V H T G D T T G M G S I R L L E A V R L S R
121 V D C R F Y Q A S S S E M F G A S P P P Q N E S T P F Y P R
151 S P Y G A A K V F S Y W T T R N Y R E A Y G L F A V N G I L
181 F N H E S P R R G E T F V T R K I T R A V A R I R A G V Q S
211 E V Y M G N L D A I R D W G Y A P E Y V E G M W R M L Q A P
241 E P D D Y V L A T G R G Y T V R E F A Q A A F D H V G L D W
271 Q K H V K F D D R Y L R P T E V D S L V G D A D R A A Q S L
301 G W K A S V H T G E L A R I M V D A D I A A S E C D G T P W
331 I D T P M L P G W G G V S
```

Seq. ID No.13

```
   1 gtgcgatggc acaccatgga tcgacacgcc gatgttgcct ggttggggca gagtaagttg
  61 acgactacac ctgggcctct ggaccgcgca acgcccgtgt atatcgccgg tcatcggggg
 121 ctggtcggct cagcgctcgt acgtagattt gaggccgagg ggttcaccaa tctcattgtg
 181 cgatcacgcg atgagattga tctgacggac cgagccgcaa cgtttgattt tgtgtctgag
 241 acaagaccac aggtgatcat cgatgcggcc gcacgggtcg gcggcatcat ggcgaataac
 301 acctatcccg cggacttctt gtccgaaaac ctccgaatcc agaccaattt gctcgacgca
 361 gctgtcgccg tgcgtgtgcc gcggctcctt ttcctcggtt cgtcatgcat ctacccgaag
 421 tacgctccgc aacctatcca cgagagtgct ttattgactg gccctttgga gcccaccaac
 481 gacgcgtatg cgatcgccaa gatcgccggt atcctgcaag ttcaggcggt taggcgccaa
 541 tatgggctgg cgtggatctc tgcgatgccg actaacctct acggacccgg cgacaacttc
 601 tccccgtccg ggtcgcatct cttgccggcg ctcatccgtc gatatgagga agccaaagct
 661 ggtggtgcag aagaggtgac gaattggggg accggtactc cgcggcgcga acttctgcat
 721 gtcgacgatc tggcgagcgc atgcctgttc cttttggaac atttcgatgg tccgaaccac
 781 gtcaacgtgg gcaccggcgt cgatcacagc attagcgaga tcgcagacat ggtcgctaca
 841 gcggtgggct acatcggcga aacacgttgg gatccaacta aacccgatgg aaccccgcgc
 901 aaactattgg acgtctccgc gctacgcgag ttgggttggc gcccgcgaat cgcactgaaa
 961 gacggcatcg atgcaacggt gtcgtggtac cgcacaaatg ccgatgccgt gaggaggtaa
```

Seq. ID No.14

```
   1  V R W H T M D R H A D V A W L G Q S K L T T T P G P L D R A
  31  T P V Y I A G H R G L V G S A L V R R F E A E G F T N L I V
  61  R S R D E I D L T D R A A T F D F V S E T R P Q V I I D A A
  91  A R V G G I M A N N T Y P A D F L S E N L R I Q T N L L D A
 121  A V A V R V P R L L F L G S S C I Y P K Y A P Q P I H E S A
 151  L L T G P L E P T N D A Y A I A K I A G I L Q V Q A V R R Q
 181  Y G L A W I S A M P T N L Y G P G D N F S P S G S H L L P A
 211  L I R R Y E E A K A G G A E E V T N W G T G T P R R E L L H
 241  V D D L A S A C L F L L E H F D G P N H V N V G T G V D H S
 271  I S E I A D M V A T A V G Y I G E T R W D P T K P D G T P R
 301  K L L D V S A L R E L G W R P R I A L K D G I D A T V S W Y
 331  R T N A D A V R R
```

Seq. ID No.15

```
   1 gtgcgatggc acaccatgga tcgacacgcc gatgttgcct ggttggggcg gagtaagttg
  61 acgactacac ctgggcctct ggaccgcgca acgcccgtgt atatcgccgg tcatcggggg
 121 ctggtcggct cagcgctcgt acgtagattt gaggccgagg ggttcaccaa tctcattgtg
 181 cgatcacgcg atgagattga tctgacggac cgagccgcaa cgtttgattt tgtgtctgag
 241 acaagaccac aggtgatcat cgatgcggcc gcacgggtcg gcggcatcat ggcgaataac
 301 acctatcccg cggacttctt gtccgaaaac ctccgaatcc agaccaattt gctcgacgca
 361 gctgtcgccg tgcgtgtgcc gcggctcctt ttcctcggtt cgtcatgcat ctacccgaag
 421 tacgctccgc aacctatcca cgagagtgct ttattgactg gccctttgga gcccaccaac
 481 gacgcgtatg cgatcgccaa gatcgccggt atcctgcaag ttcaggcggt taggcgccaa
 541 tatgggctgg cgtggatctc tgcgatgccg actaacctct acggacccgg cgacaacttc
 601 tccccgtccg ggtcgcatct cttgccggcg ctcatccgtc gatatgagga agccaaagct
 661 ggtggtgcag aagaggtgac gaattggggg accggtactc cgcggcgcga acttctgcat
 721 gtcgacgatc tggcgagcgc atgcctgttc cttttggaac atttcgatgg tccgaaccac
 781 gtcaacgtgg gcaccggcgt cgatcacagc attagcgaga tcgcagacat ggtcgctacg
 841 gcggtgggct acatcggcga aacacgttgg gatccaacta aacccgatgg aaccccgcgc
 901 aaactattgg acgtctccgc gctacgcgag ttgggttggc gcccgcgaat cgcactgaaa
 961 gacggcatcg atgcaacggt gtcgtggtac cgcacaaatg ccgatgccgt gaggaggtaa
```

Seq. ID No.16

```
   1  V R W H T M D R H A D V A W L G R S K L T T T P G P L D R A
  31  T P V Y I A G H R G L V G S A L V R R F E A E G F T N L I V
  61  R S R D E I D L T D R A A T F D F V S E T R P Q V I I D A A
  91  A R V G G I M A N N T Y P A D F L S E N L R I Q T N L L D A
 121  A V A V R V P R L L F L G S S C I Y P K Y A P Q P I H E S A
 151  L L T G P L E P T N D A Y A I A K I A G I L Q V Q A V R R Q
 181  Y G L A W I S A M P T N L Y G P G D N F S P S G S H L L P A
 211  L I R R Y E E A K A G G A E E V T N W G T G T P R R E L L H
 241  V D D L A S A C L F L L E H F D G P N H V N V G T G V D H S
 271  I S E I A D M V A T A V G Y I G E T R W D P T K P D G T P R
 301  K L L D V S A L R E L G W R P R I A L K D G I D A T V S W Y
 331  R T N A D A V R R
```

Seq. ID No.17

```
   1 atggattttt tgcgcaacgc cggcttgatg gctcgtaacg ttagtaccga gatgctgcgc
  61 cacttcgaac gaaagcgcct attagtaaac caattcaaag catacggagt caacgttgtt
 121 attgatgtcg gtgctaactc cggccagttc ggtagcgctt tgcgtcgtgc aggattcaag
 181 agccgtatcg tttcctttga acctctttcg gggccatttg cgcaactaac gcgcaagtcg
 241 gcatcggatc cactatggga gtgtcaccag tatgccctag gcgacgccga tgagacgatt
 301 accatcaatg tggcaggcaa tgcgggggca agtagttccg tgctgccgat gcttaaaagt
 361 catcaagatg cctttcctcc cgcgaattat attggcaccg aagacgttgc aatacaccgc
 421 cttgattcgg ttgcatcaga atttctgaac cctaccgatg ttacttttcct gaagatcgac
 481 gtacagggtt tcgagaagca ggttatcacg ggcagtaagt caacgcttaa cgaaagctgc
 541 gtcggcatgc aactcgaact ttcttttatt ccgttgtacg aaggtgacat gctgattcat
 601 gaagcgcttg aacttgtcta ttccctaggt ttcagactga cgggtttgtt gcccggcttt
 661 acggatccgc gcaatggtcg aatgcttcaa gctgacggca ttttcttccg tggggacgat
 721 tga
```

Seq. ID No.18

```
  1  M D F L R N A G L M A R N V S T E M L R H F E R K R L L V N
 31  Q F K A Y G V N V V I D V G A N S G Q F G S A L R R A G F K
 61  S R I V S F E P L S G P F A Q L T R K S A S D P L W E C H Q
 91  Y A L G D A D E T I T I N V A G N A G A S S S V L P M L K S
121  H Q D A F P P A N Y I G T E D V A I H R L D S V A S E F L N
151  P T D V T F L K I D V Q G F E K Q V I T G S K S T L N E S C
181  V G M Q L E L S F I P L Y E G D M L I H E A L E L V Y S L G
211  F R L T G L L P G F T D P R N G R M L Q A D G I F F R G D D
```

Seq. ID No.19

```
   1 atggattttt tgcgcaacgc cggcttgatg gctcgtaacg ttagcaccga gatgctgcgc
  61 cacttcgaac gaaagcgcct attagtaaac caattcaaag catacggagt caacgttgtt
 121 attgatgtcg gtgctaactc cggccagttc ggtagcgctt tgcgtcgtgc aggattcaag
 181 agccgtatcg tttcctttga acctctttcg gggccatttg cgcaactaac gcgcgagtcg
 241 gcatcggatc cactatggga gtgtcaccag tatgccctag gcgacgccga tgagacgatt
 301 accatcaatg tggcaggcaa tgcgggggca agtagttccg tgctgccgat gcttaaaagt
 361 catcaagatg cctttcctcc cgcgaattat attggcaccg aagacgttgc aatacaccgc
 421 cttgattcgg ttgcatcaga atttctgaac cctaccgatg ttactttcct gaagatcgac
 481 gtacagggtt tcgagaagca ggttatcgcg ggcagtaagt caacgcttaa cgaaagctgc
 541 gtcggcatgc aactcgaact ttcttttatt ccgttgtacg aaggtgacat gctgattcat
 601 gaagcgcttg aacttgtcta ttccctaggt ttcagactga cgggtttgtt gcccggattt
 661 acggatccgc gcaatggtcg aatgcttcaa gctgacggca ttttcttccg tggggacgat
 721 tga
```

Seq. ID No.20

```
  1  M D F L R N A G L M A R N V S T E M L R H F E R K R L L V N
 31  Q F K A Y G V N V V I D V G A N S G Q F G S A L R R A G F K
 61  S R I V S F E P L S G P F A Q L T R E S A S D P L W E C H Q
 91  Y A L G D A D E T I T I N V A G N A G A S S S V L P M L K S
121  H Q D A F P P A N Y I G T E D V A I H R L D S V A S E F L N
151  P T D V T F L K I D V Q G F E K Q V I A G S K S T L N E S C
181  V G M Q L E L S F I P L Y E G D M L I H E A L E L V Y S L G
211  F R L T G L L P G F T D P R N G R M L Q A D G I F F R G D D
```

Seq. ID No.21

```
   1 atgactgcgc cagtgttctc gataattatc cctaccttca atgcagcggt gacgctgcaa
  61 gcctgcctcg gaagcatcgt cgggcagacc taccgggaag tggaagtggt ccttgtcgac
 121 ggcggttcga ccgatcggac cctcgacatc gcgaacagtt tccgcccgga actcggctcg
 181 cgactggtcg ttcacagcgg gcccgatgat ggcccctacg acgccatgaa ccgcggcgtc
 241 ggcgtggcca caggcgaatg ggtacttttt ttaggcgccg acgacaccct ctacgaacca
 301 accacgttgg cccaggtagc cgcttttctc ggcgaccatg cggcaagcca tcttgtctat
 361 ggcgatgttg tgatgcgttc gacgaaaagc cggcatgccg gacctttcga cctcgaccgc
 421 ctcctatttg agacgaattt gtgccaccaa tcgatctttt accgccgtga gcttttcgac
 481 ggcatcggcc cttacaacct gcgctaccga gtctgggcgg actgggactt caatattcgc
 541 tgcttctcca acccggcgct gattacccgc tacatggacg tcgtgatttc cgaatacaac
 601 gacatgaccg gcttcagcat gaggcagggg actgataaag agttcagaaa acggctgcca
 661 atgtacttct gggttgcagg gtgggagact tgcaggcgca tgctggcgtt tttgaaagac
 721 aaggagaatc gccgtctggc cttgcgtacg cggttgataa gggttaaggc cgtctccaaa
 781 gaacgaagcg cagaaccgta g
```

Seq. ID No.22

```
   1  M T A P V F S I I I P T F N A A V T L Q A C L G S I V G Q T
  31  Y R E V E V V L V D G G S T D R T L D I A N S F R P E L G S
  61  R L V V H S G P D D G P Y D A M N R G V G V A T G E W V L F
  91  L G A D D T L Y E P T T L A Q V A A F L G D H A A S H L V Y
 121  G D V V M R S T K S R H A G P F D L D R L L F E T N L C H Q
 151  S I F Y R R E L F D G I G P Y N L R Y R V W A D W D F N I R
 181  C F S N P A L I T R Y M D V V I S E Y N D M T G F S M R Q G
 211  T D K E F R K R L P M Y F W V A G W E T C R R M L A F L K D
 241  K E N R R L A L R T R L I R V K A V S K E R S A E P
```

Seq. ID No.23

```
   1 atgactgcgc cagtgttctc gataattatc cctaccttca atgcagcggt gacgctgcaa
  61 gcctgcctcg gaagcatcgt cgggcagacc taccgggaag tggaagtggt ccttgtcgac
 121 ggcggttcga ccgatcggac cctcgacatc gcgaacagtt tccgcccgga actcggctcg
 181 cgactggtcg ttcacagcgg gcccgatgat ggcccctacg acgccatgaa ccgcggcgtc
 241 ggcgtagcca caggcgaatg ggtacttttt ttaggcgccg acgacaccct ctacgaacca
 301 accacgttgg cccaggtagc cgcttttctc ggcgaccatg cggcaagcca tcttgtctat
 361 ggcgatgttg tgatgcgttc gacgaaaagc cggcatgccg gacctttcga cctcgaccgc
 421 ctcctatttg agacgaattt gtgccaccaa tcgatctttt accgccgtga gcttttcgac
 481 ggcatcggcc cttacaacct gcgctaccga gtctgggcgg actgggactt caatattcgc
 541 tgcttctcca acccggcgct gattacccgc tacatggacg tcgtgatttc cgaatacaac
 601 gacatgaccg gcttcagcat gaggcagggg actgataaag agttcagaaa acggctgcca
 661 atgtacttct gggttgcagg gtgggagact tgcaggcgca tgctggcgtt tttgaaagac
 721 aaggagaatc gccgtctggc cttgcgtacg cggttgataa gggttaaggc cgtctccaaa
 781 gaacgaagcg cagaaccgta g
```

Seq. ID No.24

```
  1 M T A P V F S I I I P T F N A A V T L Q A C L G S I V G Q T
 31 Y R E V E V V L V D G G S T D R T L D I A N S F R P E L G S
 61 R L V V H S G P D D G P Y D A M N R G V G V A T G E W V L F
 91 L G A D D T L Y E P T T L A Q V A A F L G D H A A S H L V Y
121 G D V V M R S T K S R H A G P F D L D R L L F E T N L C H Q
151 S I F Y R R E L F D G I G P Y N L R Y R V W A D W D F N I R
181 C F S N P A L I T R Y M D V V I S E Y N D M T G F S M R Q G
211 T D K E F R K R L P M Y F W V A G W E T C R R M L A F L K D
241 K E N R R L A L R T R L I R V K A V S K E R S A E P
```

Seq. ID No.25

```
   1 gtggccagca gaagtcccca ctccgctgcg ggtggttggc taattcttgg cggctccctt
  61 cttgtggtcg gcgtggcgca tccggtagga ctcgccggag gtgacgacga tgctggcgtg
 121 gtgcagcagc cgatcgagga tgctggcggc ggtggtgtgc tcgggcagga atcgccccca
 181 ttgttcgaag ggccaatgcg aggcgatggc cagggagcgg cgctcgtagc cggcagccac
 241 gagccggaac aacagttgag tcccggtgtc gtcgagcggg gcgaagccga tctcgtccaa
 301 gatgaccaga tccgcgcgga gcagggtgtc gatgatcttg ccgacggtgt tgtcggccag
 361 gccgcggtag aggacctcga tcaggtcggc ggcggtgaag tagcggactt tgaatccggc
 421 gtggacggca gcgtgcccgc agccgatgag caggtgactt ttgcccgtac caggtgggcc
 481 aatgaccgcc aggttctgtt gtgcccgaat ccattccagg ctcgacaggt agtcgaacgt
 541 ggctgcggtg atcgacgatc cggtgacgtc gaacccgtcg agggtcttgg tgaccgggaa
 601 ggctgcggcc ttgagacggt tggcggtgtt ggaggcatcg cgggcagcga tctcggcctc
 661 aaccaacgtc cgcaggatct cctccggtgt ccagcgttgc gtcttggcga cttgcaacac
 721 ctcggcggcg ttgcggcgca ccgtggccag cttcaaccgc cgcagcgccg cgtcaaggtc
 781 agcagccagc ggtgccgccg aggacggtgc caccggcttg gcagcggtgg tcatgaggcc
 841 gtcccgtcgg tggtgttgat cttgtag
```

Seq. ID No.26

```
  1 V A S R S P H S A A G G W L I L G G S L L V V G V A H P V G
 31 L A G G D D D A G V V Q Q P I E D A G G G G V L G Q E S P P
 61 L F E G P M R G D G Q G A A L V A G S H E P E Q Q L S P G V
 91 V E R G E A D L V Q D D Q I R A E Q G V D D L A D G V V G Q
121 A A V E D L D Q V G G G E V A D F E S G V D G S V P A A D E
151 Q V T F A R T R W A N D R Q V L L C P N P F Q A R Q V V E R
181 G C G D R R S G D V E P V E G L G D R E G C G L E T V G G V
211 G G I A G S D L G L N Q R P Q D L L R C P A L R L G D L Q H
241 L G G V A A H R G Q L Q P P Q R R V K V S S Q R C R R G R C
271 H R L G S G G H E A V P S V V L I L
```

Seq. ID No.27

```
   1 atgggctgcc tcaaaggtgg tgtcgtcgcc aatgttgttg ttccaacacc ggattatgtg
  61 cgattcgcgt cccactatgg cttcgttccg gacttctgcc acggtgcgga tccgcaatcg
 121 aagggcatcg tggagaacct ctgtggctac gctcaggacg accttgcggt gccgctgctg
 181 accgaagctg cgttagccgg tgagcaggtc gacctacgtg ccctcaacgc ccaggcgcaa
 241 ctatggtgcg ccgaggtcaa tgccacggtc cactcggaga tctgcgccgt gcccaacgat
 301 cgcttggttg acgagcgcac cgtcttgagg gagctgccct cgctgcggcc gacgatcggc
 361 tcggggtcgg tgcgccgtaa ggtcgacggc ctctcgtgca tccgttacgg ctcagctcgt
 421 tactcggtgc ctcagcggct cgtcggtgcc accgtggcgg tggtggtcga tcatggcgcc
 481 ctgatcctgt tggaacctgc gaccggtgtg atcgtggccg agcacgagct cgtcagccca
 541 ggtgaggtgt ccatcctcga tgaacactac gacggaccca gacccgcacc ctcgcgtggt
 601 cctcgcccga aaacccaagc agagaaacga ttctgcgcat tgggaaccga agcgcagcag
 661 ttcctcgtcg gtgctgctgc gatcggcaac acccgactga aatccgaact cgacattctg
 721 ctcggccttg gcgccgccca cggcgaacag gctttgattg acgcgctgcg ccgggcggtt
 781 gcgtttcgcc ggttccgcgc tgccgacgtg cgctcgatcc tggccgccgg cgccggcacc
 841 ccacaacccc gccccgccgg cgacgcactc gtgctcgatc tgcccaccgt cgagacccgc
 901 tcgttggagg cctacaagat caacaccacc gacgggacgg cctcatgacc accgctgcca
 961 agccggtggc accgtcctcg gcggcaccgc tggctgctga ccttgacgcg gcgctgcggc
1021 ggttgaagct ggccacggtg cgccgcaacg ccgccgaggt gttgcaagtc gccaagacgc
1081 aacgctggac accggaggag atcctgcgga cgttggttga ggccgagatc gctgcccgcg
1141 atgcctccaa caccgccaac cgtctcaagg ccgcagcctt cccggtcacc aagaccctcg
1201 acgggttcga cgtcaccgga tcgtcgatca ccgcagccac gttcgactac ctgtcgagcc
1261 tggaatggat tcgggcacaa cagaacctgg cggtcattgg cccacctggt acgggcaaaa
1321 gtcacctgct catcggctgc gggcacgctg ccgtccacgc cggattcaaa gtccgctact
1381 tcaccgccgc cgacctgatc gaggtcctct accgcggcct ggccgacaac accgtcggca
1441 agatcatcga caccctgctc cgcgcggatc tggtcatctt ggacgagatc ggcttcgccc
1501 cgctcgacga caccgggact caactgttgt tccggctcgt ggctgccggc tacgagcgcc
1561 gctccctggc catcgcctcg cattggccct tcgaacaatg ggggcgattc ctgcccgagc
1621 acaccaccgc cgccagcatc ctcgatcggc tgctgcacca cgccagcatc gtcgtcacct
1681 ccggcgagtc ctaccggatg cgccacgccg accacaagaa gggagccgcc aagaattag
```

Seq. ID No.28

```
   1  M G C L K G G V V A N V V V P T P D Y V R F A S H Y G F V P
  31  D F C H G A D P Q S K G I V E N L C G Y A Q D D L A V P L L
  61  T E A A L A G E Q V D L R A L N A Q A Q L W C A E V N A T V
  91  H S E I C A V P N D R L V D E R T V L R E L P S L R P T I G
 121  S G S V R R K V D G L S C I R Y G S A R Y S V P Q R L V G A
 151  T V A V V V D H G A L I L L E P A T G V I V A E H E L V S P
 181  G E V S I L D E H Y D G P R P A P S R G P R P K T Q A E K R
 211  F C A L G T E A Q Q F L V G A A A I G N T R L K S E L D I L
 241  L G L G A A H G E Q A L I D A L R R A V A F R R F R A A D V
 271  R S I L A A G A G T P Q P R P A G D A L V L D L P T V E T R
 301  S L E A Y K I N T T D G T A S
```

Seq. ID No.29

```
  1  M T T A A K P V A P S S A A P L A A D L D A A L R R L K L A
 31  T V R R N A A E V L Q V A K T Q R W T P E E I L R T L V E A
 61  E I A A R D A S N T A N R L K A A A F P V T K T L D G F D V
 91  T G S S I T A A T F D Y L S S L E W I R A Q Q N L A V I G P
121  P G T G K S H L L I G C G H A A V H A G F K V R Y F T A A D
151  L I E V L Y R G L A D N T V G K I I D T L L R A D L V I L D
181  E I G F A P L D D T G T Q L L F R L V A A G Y E R R S L A I
211  A S H W P F E Q W G R F L P E H T T A A S I L D R L L H H A
241  S I V V T S G E S Y R M R H A D H K K G A A K N
```

Seq. ID No.30

```
   1 gtgacgtctg ctccgaccgt ctcggtgata acgatctcgt tcaacgacct cgacgggttg
  61 cagcgcacgg tgaaaagtgt gcgggcgcaa cgctaccggg gacgcatcga gcacatcgta
 121 atcgacggtg gcagcggcga cgacgtggtg gcatacctgt ccgggtgtga accaggcttc
 181 gcgtattggc agtccgagcc cgacggcggg cggtacgacg cgatgaacca gggcatcgcg
 241 cacgcatcgg gtgatctgtt gtggttcttg cactccgccg atcgtttttc cgggcccgac
 301 gtggtagccc aggccgtgga ggcgctatcc ggcaagggac cggtgtccga attgtggggc
 361 ttcgggatgg atcgtctcgt cgggctcgat cgggtgcgcg gcccgatacc tttcagcctg
 421 cgcaaattcc tggccggcaa gcaggttgtt ccgcatcaag catcgttctt cggatcatcg
 481 ctggtggcca agatcggtgg ctacgacctt gatttcggga tcgccgccga ccaggaattc
 541 atattgcggg ccgcgctggt atgcgagccg gtcacgattc ggtgtgtgct gtgcgagttc
 601 gacaccacgg gcgtcggctc gcaccgggaa ccaagcgcgg tcttcggtga tctgcgccgc
 661 atgggcgacc ttcatcgccg ctacccgttc gggggaaggc gaatatcaca tgcctaccta
 721 cgcggccggg agttctacgc ctacaacagt cgattctggg aaaacgtctt cacgcgaatg
 781 tcgaaatag
```

Seq. ID No.31

```
  1  M T S A P T V S V I T I S F N D L D G L Q R T V K S V R A Q
 31  R Y R G R I E H I V I D G G S G D D V V A Y L S G C E P G F
 61  A Y W Q S E P D G G R Y D A M N Q G I A H A S G D L L W F L
 91  H S A D R F S G P D V V A Q A V E A L S G K G P V S E L W G
121  F G M D R L V G L D R V R G P I P F S L R K F L A G K Q V V
151  P H Q A S F F G S S L V A K I G G Y D L D F G I A A D Q E F
181  I L R A A L V C E P V T I R C V L C E F D T T G V G S H R E
211  P S A V F G D L R R M G D L H R R Y P F G G R R I S H A Y L
241  R G R E F Y A Y N S R F W E N V F T R M S K
```

Seq. ID No.32

```
   1 gtgaagcgag cgctcatcac cggaatcacc ggccaggacg gctcgtatct cgccgaactg
  61 ctgctggcca aggggtatga ggttcacggg ctcatccggc gcgcttcgac gttcaacacc
 121 tcgcggatcg atcacctcta cgtcgacccg caccaaccgg gcgcgcggct gtttctgcac
 181 tatggtgacc tgatcgacgg aacccggttg gtgaccctgc tgagcaccat cgaacccgac
 241 gaggtgtaca acctggcggc gcagtcacac gtgcgggtga gcttcgacga acccgtgcac
 301 accggtgaca ccaccggcat gggatccatg cgactgctgg aagccgttcg gctctctcgg
 361 gtgcactgcc gcttctatca ggcgtcctcg tcggagatgt tcggcgcctc gccgccaccg
 421 cagaacgagc tgacgccgtt ctacccgcgg tcaccgtatg gcgccgccaa ggtctattcg
 481 tactgggcga cccgcaatta tcgcgaagcg tacggattgt tcgccgttaa cggcatcttg
 541 ttcaatcacg aatcaccgcg gcgcggtgag acgttcgtga cccgaaagat caccagggcc
 601 gtggcacgca tcaaggccgg tatccagtcc gaggtctata tgggcaatct ggatgcggtc
 661 cgcgactggg ggtacgcgcc cgaatacgtc gaaggcatgt ggcggatgct gcagaccgac
 721 gagcccgacg acttcgtttt ggcgaccggg cgcggtttca ccgtgcgtga gttcgcgcgg
 781 gccgcgttcg agcatgccgg tttggactgg cagcagtacg tgaaattcga ccaacgctat
 841 ctgcggccca ccgaggtgga ttcgctgatc ggcgacgcga ccaaggctgc cgaattgctg
 901 ggctggaggg cttcggtgca cactgacgag ttggctcgga tcatggtcga cgcggacatg
 961 gcggcgctgg agtgcgaagg caagccgtgg atcgacaagc cgatgatcgc cggccggaca
1021 tga
```

Seq. ID No.33

```
   1  M K R A L I T G I T G Q D G S Y L A E L L L A K G Y E V H G
  31  L I R R A S T F N T S R I D H L Y V D P H Q P G A R L F L H
  61  Y G D L I D G T R L V T L L S T I E P D E V Y N L A A Q S H
  91  V R V S F D E P V H T G D T T G M G S M R L L E A V R L S R
 121  V H C R F Y Q A S S S E M F G A S P P P Q N E L T P F Y P R
 151  S P Y G A A K V Y S Y W A T R N Y R E A Y G L F A V N G I L
 181  F N H E S P R R G E T F V T R K I T R A V A R I K A G I Q S
 211  E V Y M G N L D A V R D W G Y A P E Y V E G M W R M L Q T D
 241  E P D D F V L A T G R G F T V R E F A R A A F E H A G L D W
 271  Q Q Y V K F D Q R Y L R P T E V D S L I G D A T K A A E L L
 301  G W R A S V H T D E L A R I M V D A D M A A L E C E G K P W
 331  I D K P M I A G R T
```

Seq. ID No.34

```
   1 atgaggctgg cccgtcgcgc tcggaacatc ttgcgtcgca acggcatcga ggtgtcgcgc
  61 tactttgccg aactggactg ggaacgcaat ttcttgcgcc aactgcaatc gcatcgggtc
 121 agtgccgtgc tcgatgtcgg ggccaattcg gggcagtacg ccaggggtct gcgcggcgcg
 181 ggcttcgcgg gccgcatcgt ctcgttcgag ccgctgcccg ggccctttgc cgtcttgcag
 241 cgcagcgcct ccacggaccc gttgtgggaa tgccggcgct gtgcgctggg cgatgtcgat
 301 ggaaccatct cgatcaacgt cgccggcaac gagggcgcca gcagttccgt cttgccgatg
 361 ttgaaacgac atcaggacgc ctttccacca gccaactacg tgggcgccca acgggtgccg
 421 atacatcgac tcgattccgt ggctgcagac gttctgcggc ccaacgatat tgcgttcttg
 481 aagatcgacg ttcaaggatt cgagaagcag gtgatcgcgg gtggcgattc aacggtgcac
 541 gaccgatgcg tcggcatgca gctcgagctg tctttccagc cgttgtacga gggtggcatg
 601 ctcatccgcg aggcgctcga tctcgtggat tcgttgggct ttacgctctc gggattgcaa
 661 cccggtttca ccgaccccg caacggtcga atgctgcagg ccgatggcat cttcttccgg
 721 ggcagcgatt ga
```

Seq. ID No.35

```
  1  M R L A R R A R N I L R R N G I E V S R Y F A E L D W E R N
 31  F L R Q L Q S H R V S A V L D V G A N S G Q Y A R G L R G A
 61  G F A G R I V S F E P L P G P F A V L Q R S A S T D P L W E
 91  C R R C A L G D V D G T I S I N V A G N E G A S S S V L P M
121  L K R H Q D A F P P A N Y V G A Q R V P I H R L D S V A A D
151  V L R P N D I A F L K I D V Q G F E K Q V I A G G D S T V H
181  D R C V G M Q L E L S F Q P L Y E G G M L I R E A L D L V D
211  S L G F T L S G L Q P G F T D P R N G R M L Q A D G I F F R
241  G S D
```

Seq. ID No.36

```
   1 gtgaaatcgt tgaaactcgc tcgtttcatc gcgcgtagcg ccgccttcga ggtttcgcgc
  61 cgctattctg agcgagacct gaagcaccag tttgtgaagc aactcaaatc gcgtcgggta
 121 gatgtcgttt tcgatgtcgg cgccaactca ggacaatacg ccgccggcct ccgccgagca
 181 gcatataagg gccgcattgt ctcgttcgaa ccgctatccg gaccgtttac gatcttggaa
 241 agcaaagcgt caacggatcc actttgggat tgccggcagc atgcgttggg cgattctgat
 301 ggaacggtta cgatcaatat cgcaggaaac gccggtcaga gcagttccgt cttgcccatg
 361 ctgaaaagtc atcagaacgc ttttcccccg gcaaactatg tcggtaccca agaggcgtcc
 421 atacatcgac ttgattccgt ggcgccagaa tttctaggca tgaacggtgt cgcttttctc
 481 aaggtcgacg ttcaaggctt tgaaaagcag gtgctcgccg ggggcaaatc aaccatagat
 541 gaccattgcg tcggcatgca actcgaactg tccttcctgc cgttgtacga aggtggcatg
 601 ctcattcctg aagccctcga tctcgtgtat tccttgggct tcacgttgac gggattgctg
 661 ccttgtttca ttgatgcaaa taatggtcga atgttgcagg ccgacggcat ctttttccgc
 721 gaggacgatt ga
```

Seq. ID No.37

```
  1  M K S L K L A R F I A R S A A F E V S R R Y S E R D L K H Q
 31  F V K Q L K S R R V D V V F D F T V G A N S G Q Y A A G L R
 61  R A A Y K G R I V S F E P L S G P F T I L E S K A S T D P L
 91  W D C R Q H A L G D S D G T V T I N I A G N A G Q S S S V L
121  P M L K S H Q N A F P P A N Y V G T Q E A S I H R L D S V A
151  P E F L G M N G V A F L K V D V Q G F E K Q V L A G G K S T
181  I D D H C V G M Q L E L S F L P L Y E G G M L I P E A L D L
211  V Y S L G F T L T G L L P C F I D A N N G R M L Q A D G I F
241  F R E D D
```

Seq. ID No.38

```
   1 atggtgcaga cgaaacgata cgccggcttg accgcagcta acacaaagaa agtcgccatg
  61 gccgcaccaa tgttttcgat catcatcccc accttgaacg tggctgcggt attgcctgcc
 121 tgcctcgaca gcatcgcccg tcagacctgc ggtgacttcg agctggtact ggtcgacggc
 181 ggctcgacgg acgaaaccct cgacatcgcc aacattttcg cccccaacct cggcgagcgg
 241 ttgatcattc atcgcgacac cgaccagggc gtctacgacg ccatgaaccg cggcgtggac
 301 ctggccaccg gaacgtggtt gctctttctg ggcgcggacg acagcctgta cgaggctgac
 361 accctggcgc gggtggccgc cttcattggc gaacacgagc ccagcgatct ggtatatggc
 421 gacgtgatca tgcgctcaac caatttccgc tggggtggcg ccttcgacct cgaccgtctg
 481 ttgttcaagc gcaacatctg ccatcaggcg atcttctacc gccgcggact cttcggcacc
 541 atcggtccct acaacctccg ctaccgggtc ctggccgact gggacttcaa tattcgctgc
 601 ttttccaacc cagcgctcgt cacccgctac atgcacgtgg tcgttgcaag ctacaacgaa
 661 ttcggcgggc tcagcaatac gatcgtcgac aaggagtttt tgaagcggct gccgatgtcc
 721 acgagactcg gcataaggct ggtcatagtt ctggtgcgca ggtggccaaa ggtgatcagc
 781 agggccatgg taatgcgcac cgtcatttct tggcggcgcc gacgttag
```

Seq. ID No.39

```
   1  M V Q T K R Y A G L T A A N T K K V A M A A P M F S I I I P
  31  T L N V A A V L P A C L D S I A R Q T C G D F E L V L V D G
  61  G S T D E T L D I A N I F A P N L G E R L I I H R D T D Q G
  91  V Y D A M N R G V D L A T G T W L L F L G A D D S L Y E A D
 121  T L A R V A A F I G E H E P S D L V Y G D V I M R S T N F R
 151  W G G A F D L D R L L F K R N I C H Q A I F Y R R G L F G T
 181  I G P Y N L R Y R V L A D W D F N I R C F S N P A L V T R Y
 211  M H V V V A S Y N E F G G L S N T I V D K E F L K R L P M S
 241  T R L G I R L V I V L V R R W P K V I S R A M V M R T V I S
 271  W R R R R
```

Seq 40:

GATGCCGTGAGGAGGTAAAGCTGC

Seq 41:

GATACGGCTCTTGAATCCTGCACG

ART 34

CLAIMS

1. A polypeptide in substantially isolated form which comprises a sequence selected from the sequences of Seq.ID.No: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28 and 29, or a polypeptide substantially homologous thereto.

2. A polypeptide in substantially isolated form which comprises a sequence selected from the sequences of Seq.ID.No: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28 and 29.

3. A polypeptide which comprises a fragment of a polypeptide defined in claim 1 or 2, said fragment comprising at least 12 amino acids and an epitope.

4. A polynucleotide in substantially isolated form which encodes a polypeptide according to any one of claims 1 to 3.

5. A polynucleotide in substantially isolated form which is capable of selectively hybridizing to Seq.ID.No: 3 or 4 or a fragment thereof.

6. A polynucleotide fragment according to claim 5 which comprises a sequence selected from the sequences of Seq.ID.No: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25 and 27, or a polynucleotide at least 90% homologous thereto.

7. A polynucleotide in substantially isolated form comprising a sequence selected from the sequences of Seq.ID.No: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25 and 27.

8. A polynucleotide probe which comprises a fragment of at least 15 nucleotides of a polynucleotide as defined in any one of claims 4 to 7, optionally carrying a revealing label.

9. A recombinant vector carrying a polynucleotide as defined in any one of claims 4 to 7.

10. An antibody capable of binding a polypeptide or fragment thereof as defined in any one of claims 1 to 3.

11. An antibody capable of binding a polypeptide or fragment thereof wherein the polypeptide is a polypeptide which comprises a sequence selected from the sequences of Seq.ID.No: 31, 33, 35, 37 and 39 or is a peptide substantially homogolous thereto.

12. A test kit for detecting the presence or absence of a pathogenic mycobacterium in a sample which comprises a polynucleotide according to any one of claims 4 to 8, a polypeptide according to any one of claims 1 to 3, a polypeptide which comprises a sequence selected from the sequences of Seq.ID.No: 31, 33, 35, 37 and 39 or a polypeptide substantially homogolous thereto, or an antibody according to, any one of claims 10 or 11.

13. A method of detecting the presence or absence of antibodies in an animal or human, against a pathogenic mycobacteria in a sample which comprises:

(a) providing a polypeptide according to any one of claims 1 to 3 or a polypeptide which comprises a sequence selected from the sequences of Seq.ID.No: 31, 33, 35, 37 and 39 or a polypeptide substantially homogolous thereto, which comprises an epitope;

(b) incubating a biological sample with said polypeptide under conditions which allow for the formation of an antibody-antigen complex; and

(c) determining whether antibody-antigen complex comprising said polypeptide is formed.

14. A method of detecting the presence or absence of a polypeptide according to any one of claims 1 to 3 or a polypeptide which comprises a sequence selected from the

sequences of Seq.ID.No: 31, 33, 35, 37 and 39 or a polypeptide substantially homogolous thereto in a biological sample which method which comprises:

(a) providing an antibody according to any one of claims 10 and 11;

(b) incubating a biological sample with said antibody under conditions which allow for the formation of an antibody-antigen complex; and

(c) determining whether antibody-antigen complex comprising said antibody is formed.

15. A method of detecting the presence or absence of cell mediated immune reactivity in an animal or human, to a polypeptide according to claims 1 to 3 or a polypeptide which comprises a sequence selected from the sequences of Seq.ID.No: 31, 33, 35, 37 and 39 or a polypeptide substantially homogolous thereto, which method comprises

(a) providing a polypeptide according to any one of claims 1 to 3 or a polypeptide which comprises a sequence selected from the sequences of Seq.ID.No: 31, 33, 35, 37 and 39 or a polypeptide substantially homogolous thereto, which comprises an epitope;

(b) incubating a cell sample with said polypeptide under conditions which allow for a cellular immune response such as release of cytokines or other mediator or reaction to occur; and

(c) detecting the presence of said cytokine or mediator or cellular response in the incubate.

16. A pharmaceutical composition comprising a polypeptide according to any one of claims 1 to 3 in a suitable carrier or diluent.

17. A composition according to claim 16 or a composition comprising a polypeptide which comprises a sequence selected from the sequences of Seq.ID.No: 31, 33, 35, 37 and 39 or a polypeptide substantially homogolous thereto,

for use in the treatment or prevention of diseases caused by mycobacteria.

18. A method of treating or preventing mycobacterial disease in an animal or human caused by mycobacteria which express a polypeptide according to claims 1 to 3 or a polypeptide which comprises a sequence selected from the sequences of Seq.ID.No: 31, 33, 35, 37 and 39 or a polypeptide substantially homogolous thereto, which method comprises vaccinating or treating an animal or human with an effective amount of said polypeptide.

19. A method of treating or preventing mycobacterial diseases in animals or humans caused by mycobacteria containing the polynucleotide of Seq.ID.No: 3 or 4, which method comprises vaccinating or treating an animal or human with an effective amount of a polynucleotide according to claims 4 to 7, a vector according to claim 9 or a polynucleotide which encodes a polypeptide which comprises a sequence selected from the sequences of Seq.ID.No: 31, 33, 35, 37 and 39 or a polypeptide substantially homogolous thereto.

20. A method according to claims 18 or 19 for increasing the in vivo susceptibility of mycobacteria to antimicrobial drugs.

21. A normally pathogenic mycobacterium, whose pathogenicity is mediated in all or in part by the presence or the expression of a polypeptide as defined in any one of claims 1 to 3 or a polypeptide which comprises a sequence selected from the sequences of Seq.ID.No: 31, 33, 35, 37 and 39 or a polypeptide substantially homogolous thereto, which mycobacterium harbours an attenuating mutation in a gene encoding one of the said polypeptides.

22. A vaccine comprising a mycobacterium as claimed in claim 21.

23. A vaccine according to claim 22 wherein the mycobacteria is selected from *Mavs*, *Mptb* and *Mtb*.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: St George's Hospital Medical School
(B) STREET: Cranmer Terrace
(C) CITY: London
(E) COUNTRY: United Kingdom
(F) POSTAL CODE (ZIP): SW17 ORE

(ii) TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES IN PATHOGENIC MYCOBACTERIA AND THEIR USE AS DIAGNOSTICS, VACCINES AND TARGETS FOR CHEMOTHERAPY

(iii) NUMBER OF SEQUENCES: 41

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(v) CURRENT APPLICATION DATA:
APPLICATION NUMBER: WO PCT/GB96/03221

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 674 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GATCCAACTA AACCCGATGG AACCCCGCGC AAACTATTGG ACGTCTCCGC GCTACGCAGT 60

TGGGTTGGCG CCCGCGAATC GCACTGAAAG AGGGCATCGA TGCAACGGTG TCGTGGTACC 120

GCACAAATGC CGATGCCGTG AGGAGGTAAA GCTGCGGGCC GGCCGATGTT ATCCCTCCGG 180

CCGGACGGGT AGGGCGACCT GCCATCGAGT GGTACGGCAG TCGCCTGGCC GGCGAGGCGC 240

ATGGCCTATG TGAGTATCCC ATAGCCTGGC TTGGCTCGCC CCTACGCATT ATCAGTTGAC 300

CGCTTTCGCG CCACGTCGCA GGCTTGCGGC AGCATCCCGT TCAGGTCTCC TCATGGTCCG 360

GTGTGGCACG ACCACGCAAG CTCGAACCGA CTCGTTTCCC AATTTCGCAT GCTAATATCG 420

CTCGATGGAT TTTTTGCGCA ACGCCGGCTT GATGGCTCGT AACGTTAGCA CCGAGATGCT 480

GCGCCACTCC GAACGAAAGC GCCTATTAGT AAACCAAGTC GAAGCATACG GAGTCAACGT 540

TGTTATTGAT GTCGGTGCTA ACTCCGGCCA GTTCGGTAGC GCTTTGCGTC GTGCAGGATT 600

CAAGAGCCGT ATCGTTTCCT TTGAACCTCT TTCGGGGCCA TTTGCGCAAC TAACGCGCAA 660

GTCGGCATCG GATC 674

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 674 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

GATCCGATGC CGACTTGCGC GTTAGTTGCG CAAATGGCCC CGAAAGAGGT TCAAAGGAAA 60

CGATACGGCT CTTGAATCCT GCACGACGCA AAGCGCTACC GAACTGGCCG GAGTTAGCAC 120

CGACATCAAT AACAACGTTG ACTCCGTATG CTTCGACTTG GTTTACTAAT AGGCGCTTTC 180

GTTCGGAGTG GCGCAGCATC TCGGTGCTAA CGTTACGAGC CATCAAGCCG GCGTTGCGCA 240

AAAAATCCAT CGAGCGATAT TAGCATGCGA AATTGGGAAA CGAGTCGGTT CGAGCTTGCG 300

TGGTCGTGCC ACACCGGACC ATGAGGAGAC CTGAACGGGA TGCTGCCGCA AGCCTGCGAC 360

GTGGCGCGAA AGCGGTCAAC TGATAATGCG TAGGGGCGAG CCAAGCCAGG CTATGGGATA 420

CTCACATAGG CCATGCGCCT CGCCGGCCAG GCGACTGCCG TACCACTCGA TGGCAGGTCG 480

CCCTACCCGT CCGGCCGGAG GGATAACATC GGCCGGCCCG CAGCTTTACC TCCTCACGGC 540

ATCGGCATTT GTGCGGTACC ACGACACCGT TGCATCGATG CCCTCTTTCA GTGCGATTCG 600

CGGGCGCCAA CCCAACTGCG TAGCGCGGAG ACGTCCAATA GTTTGCGCGG GGTTCCATCG 660

GGTTTAGTTG GATC 674

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7995 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GAATTCTGGG TTGGAGACGA CGTCGAACTC CTGGTCGGTC TTGCTTCGAA TGATCGCTGT 60

GATCTGGTCG GCGGTGCCGA CAGGAACCGT CGACTTGTCG ACGATCACCT TGTACCGGTC 120

GATGTATGAC CCAATGTCGT CCGCAACCGA GAAGACGTAC GTCAGGTCCG CCGCCCCGCT 180

TTCACCCATG GGCGTCGGGA CGGCGATGAA AATGACGTCC GCGTGCTCGA TTCCGCGTTG 240

CCGGTCGGTG GTGAAGTCAA TCAGCCCGTT CTCACGGTTC CTCGCAATCA ACTCCCAACC 300

CGGGCTCGAA AATCGGGACA CTGCCTGCGA GGAGCAAATC GATCTTGGCC TGATCGATAT 360

CGACACAGAC GACATCGTTG CCGCTATCCG CGAGACAGGC GCCCGTGACG AGGCCTACAT 420

AGCCTGATCC GACCACCGAA ATTTTCAAGA TGACCCCTTC AAGTCCCCGA TCGGTCGACG 480

ACCATACTGC CGCAACTCTG TACCCTCCGT GGGTAATTCG CATGTCGCGT TCGTAAGGAG 540

CAGCCAGCGA GTCGGGGACG TTCGGTGAGA GAGTCGCAGG ACTACGAGGT TGCCGGTGCG 600

ATACATCACA GTGTTGCGTC TGTCGGCAAC GATGCAGCAA GAACCCACGG GGCAGCCCTG 660

AACTGCGCGC ATGACCGGTC CTTGTCCTGG CACCTTTGAT CGGCCACCGC TTCCATGCGA 720

ACATGACCGG AATCCATAGC GCGTGGTCAA GCAGCGGGGA GGTAGACGTC GGTGTCATCT 780

GCTCCAACCG TGTCGGTGAT AACGATTTCG CTGAACGATC TCGAGGGATT GAAAAGCACC 840

GTGGAGAGCG TTCGCGCGCA GCGCTATGGG GGGCGAATCG AGCACATCGT CATCGACGGT 900

GGATCGGGCG ACGCCGTCGT GGAGTATCTG TCCGGCGATC CTGGCTTTGC ATATTGGCAA 960

TCTCAGCCCG ACAACGGGAG ATATGACGCG ATGAATCAGG GCATTGCCCA TTCGTCGGGC 1020

GACCTGTTGT GGTTTATGCA CTCCACGGAT CGTTTCTCCG ATCCAGATGC AGTCGCTTCC 1080

GTGGTGGAGG CGCTCTCGGG GCATGGACCA GTACGTGATT TGTGGGGTTA CGGGAAAAAC 1140

AACCTTGTCG GACTCGACGG CAAACCACTT TTCCCTCGGC CGTACGGCTA TATGCCGTTT 1200

AAGATGCGGA AATTTCTGCT CGGCGCGACG GTTGCGCATC AGGCGACATT CTTCGGCGCG 1260

TCGCTGGTAG CCAAGTTGGG CGGTTACGAT CTTGATTTTG GACTCGAGGC GGACCAGCTG 1320

TTCATCTACC GTGCCGCACT AATACGGCCT CCCGTCACGA TCGACCGCGT GGTTTGCGAC 1380

TTCGATGTCA CGGGACCTGG TTCAACCCAG CCCATCCGTG AGCACTATCG GACCCTGCGG 1440

CGGCTCTGGG ACCTGCATGG CGACTACCCG CTGGGTGGGC GCAGAGTGTC GTGGGCTTAC 1500

TTGCGTGTGA AGGAGTACTT GATTCGGGCC GACCTGGCCG CATTCAACGC GGTAAAGTTC 1560

TTGCGAGCGA AGTTCGCCAG AGCTTCGCGG AAGCAAAATT CATAGAAACC AACTTCTACT 1620

GCCTGACCTG AGCAGCGCCG AGGCGCGCAG CGCGATCAGT GCGACCTGAA CGGCCAGGTG 1680

GAAAGCGCCA CCGATCCCGG CACCGAGTGC CTGACGCTTC GGATCCCTTG CACCACAACG 1740

AGAGTGAGAG CGCCATGATG AGGAAATATC GGCTGGGCGG AGTCAACGCC GGAGTGACAA 1800

AAGTGAGAAC CCGGTGAAGC GAGCGCTTAT AACAGGGATC ACGGGGCAGG ATGGTTCCTA 1860

CCTCGCCGAG CTACTACTGA GCAAGGGATA CGAGGTTCAC GGGCTCGTTC GTCGAGCTTC 1920

GACGTTTAAC ACGTCGCGGA TCGATCACCT CTACGTTGAC CCACACCAAC CGGGCGCGCG 1980

CTTGTTCTTG CACTATGCAG ACCTCACTGA CGGCACCCGG TTGGTGACCC TGCTCAGCAG 2040

TATCGACCCG GATGAGGTCT ACAACCTCGC AGCGCAGTCC CATGTGCGCG TCAGCTTTGA 2100

CGAGCCAGTG CATACCGGAG ACACCACCGG CATGGGATCG ATCCGACTTC TGGAAGCAGT 2160

CCGCCTTTCT CGGGTGGACT GCCGGTTCTA TCAGGCTTCC TCGTCGGAGA TGTTCGGCGC 2220

ATCTCCGCCA CCGCAGAACG AATCGACGCC GTTCTATCCC CGTTCGCCAT ACGGCGCGGC 2280

CAAGGTCTTC TCGTACTGGA CGACTCGCAA CTATCGAGAG GCGTACGGAT TATTCGCAGT 2340

GAATGGCATC TTGTTCAACC ATGAGTCCCC CCGGCGCGGC GAGACTTTCG TGACCCGAAA 2400

GATCACGCGT GCCGTGGCGC GCATCCGAGC TGGCGTCCAA TCGGAGGTCT ATATGGGCAA 2460

CCTCGATGCG ATCCGCGACT GGGGCTACGC GCCCGAATAT GTCGAGGGGA TGTGGAGGAT 2520

GTTGCAAGCG CCTGAACCTG ATGACTACGT CCTGGCGACA GGGCGTGGTT ACACCGTACG 2580

TGAGTTCGCT CAAGCTGCTT TTGACCATGT CGGGCTCGAC TGGCAAAAGC GCGTCAAGTT 2640

TGACGACCGC TATTTGCGTC CCACCGAGGT CGATTCGCTA GTAGGAGATG CCGACAAGGC 2700

GGCCCAGTCA CTCGGCTGGA AAGCTTCGGT TCATACTGGT GAACTCGCGC GCATCATGGT 2760

GGACGCGGAC ATCGCCGCGT TGGAGTGCGA TGGCACACCA TGGATCGACA CGCCGATGTT 2820

GCCTGGTTGG GGCAGAGTAA GTTGACGACT ACACCTGGGC CTCTGGACCG CGCAACGCCC 2880

GTGTATATCG CCGGTCATCG GGGGCTGGTC GGCTCAGCGC TCGTACGTAG ATTTGAGGCC 2940

GAGGGGTTCA CCAATCTCAT TGTGCGATCA CGCGATGAGA TTGATCTGAC GGACCGAGCC 3000

GCAACGTTTG ATTTTGTGTC TGAGACAAGA CCACAGGTGA TCATCGATGC GGCCGCACGG 3060

GTCGGCGGCA TCATGGCGAA TAACACCTAT CCCGCGGACT TCTTGTCCGA AAACCTCCGA 3120

ATCCAGACCA ATTTGCTCGA CGCAGCTGTC GCCGTGCGTG TGCCGCGGCT CCTTTTCCTC 3180

GGTTCGTCAT GCATCTACCC GAAGTACGCT CCGCAACCTA TCCACGAGAG TGCTTTATTG 3240

ACTGGCCCTT TGGAGCCCAC CAACGACGCG TATGCGATCG CCAAGATCGC CGGTATCCTG 3300

CAAGTTCAGG CGGTTAGGCG CCAATATGGG CTGGCGTGGA TCTCTGCGAT GCCGACTAAC 3360

CTCTACGGAC CCGGCGACAA CTTCTCCCCG TCCGGGTCGC ATCTCTTGCC GGCGCTCATC 3420

CGTCGATATG AGGAAGCCAA AGCTGGTGGT GCAGAAGAGG TGACGAATTG GGGGACCGGT 3480

ACTCCGCGGC GCGAACTTCT GCATGTCGAC GATCTGGCGA GCGCATGCCT GTTCCTTTTG 3540

GAACATTTCG ATGGTCCGAA CCACGTCAAC GTGGGCACCG GCGTCGATCA CAGCATTAGC 3600

GAGATCGCAG ACATGGTCGC TACAGCGGTG GGCTACATCG GCGAAACACG TTGGGATCCA 3660

ACTAAACCCG ATGGAACCCC GCGCAAACTA TTGGACGTCT CCGCGCTACG CGAGTTGGGT 3720

TGGCGCCCGC GAATCGCACT GAAAGACGGC ATCGATGCAA CGGTGTCGTG GTACCGCACA 3780

AATGCCGATG CCGTGAGGAG GTAAAGCTGC GGGTCGGCCG ATGTTATCCC TCCGGCCGGA 3840

CGGGTGGGGC GACCTGCCGT CGAGTGGTAC GGCAGTCGCC TGGCCGGCGA GGCGCGTGGC 3900

CTATGGGAGT ATCCAATAGC CTGGCTTGGC TCGCCCCTAC GCATTATCAG TTGACCGCTT 3960

TCGCGCCAGC TCGCAGGCTT GCGGCAGCAT CCCGTTCAGG TCTCCTCATG GTCCGGTGTG 4020

GCACGACCAC GCAAGCTCGA ACCGACTCGT TTCCCAATTT CGCATGCTAA TATCGCTCGA 4080

TGGATTTTTT GCGCAACGCC GGCTTGATGG CTCGTAACGT TAGTACCGAG ATGCTGCGCC 4140

ACTTCGAACG AAAGCGCCTA TTAGTAAACC AATTCAAAGC ATACGGAGTC AACGTTGTTA 4200

TTGATGTCGG TGCTAACTCC GGCCAGTTCG GTAGCGCTTT GCGTCGTGCA GGATTCAAGA 4260

GCCGTATCGT TTCCTTTGAA CCTCTTTCGG GGCCATTTGC GCAACTAACG CGCAAGTCGG 4320

CATCGGATCC ACTATGGGAG TGTCACCAGT ATGCCCTAGG CGACGCCGAT GAGACGATTA 4380

CCATCAATGT GGCAGGCAAT GCGGGGGCAA GTAGTTCCGT GCTGCCGATG CTTAAAAGTC 4440

ATCAAGATGC CTTTCCTCCC GCGAATTATA TTGGCACCGA AGACGTTGCA ATACACCGCC 4500

TTGATTCGGT TGCATCAGAA TTTCTGAACC CTACCGATGT TACTTTCCTG AAGATCGACG 4560

TACAGGGTTT CGAGAAGCAG GTTATCACGG GCAGTAAGTC AACGCTTAAC GAAAGCTGCG 4620

TCGGCATGCA ACTCGAACTT TCTTTTATTC CGTTGTACGA AGGTGACATG CTGATTCATG 4680

AAGCGCTTGA ACTTGTCTAT TCCCTAGGTT TCAGACTGAC GGGTTTGTTG CCCGGCTTTA 4740

CGGATCCGCG CAATGGTCGA ATGCTTCAAG CTGACGGCAT TTTCTTCCGT GGGGACGATT 4800

GACATAAATG CTCCGTCGGC ACCCTGCCGG TATCCAAACG GGCGATCTGG TGAGCCGGCC 4860

TCCCGGGCAC CTAATCGACT ATCTAAATTG AGGCGGCCGC GACGTGCGGC ACGAACAGGT 4920

GGCCGGCTGC TAGCGTTACA CACGTCATGA CTGCGCCAGT GTTCTCGATA ATTATCCCTA 4980

CCTTCAATGC AGCGGTGACG CTGCAAGCCT GCCTCGGAAG CATCGTCGGG CAGACCTACC 5040

GGGAAGTGGA AGTGGTCCTT GTCGACGGCG GTTCGACCGA TCGGACCCTC GACATCGCGA 5100

ACAGTTTCCG CCCGGAACTC GGCTCGCGAC TGGTCGTTCA CAGCGGGCCC GATGATGGCC 5160

CCTACGACGC CATGAACCGC GGCGTCGGCG TGGCCACAGG CGAATGGGTA CTTTTTTTAG 5220

GCGCCGACGA CACCCTCTAC GAACCAACCA CGTTGGCCCA GGTAGCCGCT TTTCTCGGCG 5280

ACCATGCGGC AAGCCATCTT GTCTATGGCG ATGTTGTGAT GCGTTCGACG AAAAGCCGGC 5340

ATGCCGGACC TTTCGACCTC GACCGCCTCC TATTTGAGAC GAATTTGTGC CACCAATCGA 5400

TCTTTTACCG CCGTGAGCTT TTCGACGGCA TCGGCCCTTA CAACCTGCGC TACCGAGTCT 5460

GGGCGGACTG GGACTTCAAT ATTCGCTGCT TCTCCAACCC GGCGCTGATT ACCCGCTACA 5520

TGGACGTCGT GATTTCCGAA TACAACGACA TGACCGGCTT CAGCATGAGG CAGGGGACTG 5580

ATAAAGAGTT CAGAAAACGG CTGCCAATGT ACTTCTGGGT TGCAGGGTGG GAGACTTGCA 5640

GGCGCATGCT GGCGTTTTTG AAAGACAAGG AGAATCGCCG TCTGGCCTTG CGTACGCGGT 5700

TGATAAGGGT TAAGGCCGTC TCCAAAGAAC GAAGCGCAGA ACCGTAGTCG CGGATCCACA 5760

TTGGACTTCT TTAACGCGTT TGCGTCCTGA TCCACCTTTC AAGCCCGTTC CGCGTAACGC 5820

GGCGCGCAGA GAGTGGTCGC ATATCGCATC ACTGTTCTCG TGCCAGTGCT TGGAAAGCGT 5880

CGAGCACTCT GGTTCGCGTT CTTGACGTTC GCGCCCGCTC CTAGAGGTAG CGTGTCACGT 5940

GACTGAAGCC AATGAGTGCA ACTCGGCGTC GCGAAAGGTT TCAGTCGCGG TTGAGCAAGA 6000

CACCGCAAGA CTACTGGAGT GCGTGCACAA GCGCCTCCAG CTCGCGGCTG AAAGCGGATG 6060

CAAAGGGATT CGAAGCTTGA GCAACATGCG AAGGGGAGAA CGGCCTATGA GGCTGGGACA 6120

GGTTTTCGAT CCGCGCGCGA ATGCACTGTC AATGGCCAAG TAGAAGTCCC CGCTGGTGGC 6180

CAGCAGAAGT CCCCACTCCG CTGCGGGTGG TTGGCTAATT CTTGGCGGCT CCCTTCTTGT 6240

GGTCGGCGTG GCGCATCCGG TAGGACTCGC CGGAGGTGAC GACGATGCTG GCGTGGTGCA 6300

GCAGCCGATC GAGGATGCTG GCGGCGGTGG TGTGCTCGGG CAGGAATCGC CCCCATTGTT 6360

CGAAGGGCCA ATGCGAGGCG ATGGCCAGGG AGCGGCGCTC GTAGCCGGCA GCCACGAGCC 6420

GGAACAACAG TTGAGTCCCG GTGTCGTCGA GCGGGGCGAA GCCGATCTCG TCCAAGATGA 6480

CCAGATCCGC GCGGAGCAGG GTGTCGATGA TCTTGCCGAC GGTGTTGTCG GCCAGGCCGC 6540

GGTAGAGGAC CTCGATCAGG TCGGCGGCGG TGAAGTAGCG GACTTTGAAT CCGGCGTGGA 6600

CGGCAGCGTG CCCGCAGCCG ATGAGCAGGT GACTTTTGCC CGTACCAGGT GGGCCAATGA 6660

CCGCCAGGTT CTGTTGTGCC CGAATCCATT CCAGGCTCGA CAGGTAGTCG AACGTGGCTG 6720

CGGTGATCGA CGATCCGGTG ACGTCGAACC CGTCGAGGGT CTTGGTGACC GGGAAGGCTG 6780

CGGCCTTGAG ACGGTTGGCG GTGTTGGAGG CATCGCGGGC AGCGATCTCG GCCTCAACCA 6840

ACGTCCGCAG GATCTCCTCC GGTGTCCAGC GTTGCGTCTT GGCGACTTGC AACACCTCGG 6900

CGGCGTTGCG GCGCACCGTG GCCAGCTTCA ACCGCCGCAG CGCCGCGTCA AGGTCAGCAG 6960

CCAGCGGTGC CGCCGAGGAC GGTGCCACCG GCTTGGCAGC GGTGGTCATG AGGCCGTCCC 7020

GTCGGTGGTG TTGATCTTGT AGGCCTCCAA CGAGCGGGTC TCGACGGTGG GCAGATCGAG 7080

CACGAGTGCG TCGCCGGCGG GGCGGGGTTG TGGGGTGCCG GCGCCGGCGG CCAGGATCGA 7140

GCGCACGTCG GCAGCGCGGA ACCGGCGAAA CGCAACCGCC CGGCGCAGCG CGTCAATCAA 7200

AGCCTGTTCG CCGTGGGCGG CGCCAAGGCC GAGCAGAATG TCGAGTTCGG ATTTCAGTCG 7260

GGTGTTGCCG ATCGCAGCAG CACCGACGAG GAACTGCTGC GCTTCGGTTC CCAATGCGCA 7320

GAATCGTTTC TCTGCTTGGG TTTTCGGGCG AGGACCACGC GAGGGTGCGG GTCTGGGTCC 7380

GTCGTAGTGT TCATCGAGGA TGGACACCTC ACCTGGGCTG ACGAGCTCGT GCTCGGCCAC 7440

GATCACACCG GTCGCAGGTT CCAACAGGAT CAGGGCGCCA TGATCGACCA CCACCGCCAC 7500

GGTGGCACCG ACGAGCCGCT GAGGCACCGA GTAACGAGCT GAGCCGTAAC GGATGCACGA 7560

GAGGCCGTCG ACCTTACGGC GCACCGACCC CGAGCCGATC GTCGGCCGCA GCGAGGGCAG 7620

CTCCCTCAAG ACGGTGCGCT CGTCAACCAA GCGATCGTTG GGCACGGCGC AGATCTCCGA 7680

GTGGACCGTG GCATTGACCT CGGCGCACCA TAGTTGCGCC TGGGCGTTGA GGGCACGTAG 7740

GTCGACCTGC TCACCGGCTA ACGCAGCTTC GGTCAGCAGC GGCACCGCAA GGTCGTCCTG 7800

AGCGTAGCCA CAGAGGTTCT CCACGATGCC CTTCGATTGC GGATCCGCAC CGTGGCAGAA 7860

GTCCGGAACG AAGCCATAGT GGGACGCGAA TCGCACATAA TCCGGTGTTG GAACAACAAC 7920

ATTGGCGACG ACACCACCTT TGAGGCAGCC CATCCGGTCG GCCAGGATCT TGGCCGGAAC 7980

CCCACCGATC GCCTC 7995

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4435 base pairs
(B) TYPE: nucleic acid

AACGCCCGTG TATATCGCCG GTCATCGGGG GCTGGTCGGC TCAGCGCTCG TACGTAGATT 1320

TGAGGCCGAG GGGTTCACCA ATCTCATTGT GCGATCACGC GATGAGATTG ATCTGACGGA 1380

CCGAGCCGCA ACGTTTGATT TTGTGTCTGA GACAAGACCA CAGGTGATCA TCGATGCGGC 1440

CGCACGGGTC GGCGGCATCA TGGCGAATAA CACCTATCCC GCGGACTTCT TGTCCGAAAA 1500

CCTCCGAATC CAGACCAATT TGCTCGACGC AGCTGTCGCC GTGCGTGTGC CGCGGCTCCT 1560

TTTCCTCGGT TCGTCATGCA TCTACCCGAA GTACGCTCCG CAACCTATCC ACGAGAGTGC 1620

TTTATTGACT GGCCCTTTGG AGCCCACCAA CGACGCGTAT GCGATCGCCA AGATCGCCGG 1680

TATCCTGCAA GTTCAGGCGG TTAGGCGCCA ATATGGGCTG GCGTGGATCT CTGCGATGCC 1740

GACTAACCTC TACGGACCCG GCGACAACTT CTCCCCGTCC GGGTCGCATC TCTTGCCGGC 1800

GCTCATCCGT CGATATGAGG AAGCCAAAGC TGGTGGTGCA GAAGAGGTGA CGAATTGGGG 1860

GACCGGTACT CCGCGGCGCG AACTTCTGCA TGTCGACGAT CTGGCGAGCG CATGCCTGTT 1920

CCTTTTGGAA CATTTCGATG GTCCGAACCA CGTCAACGTG GGCACCGGCG TCGATCACAG 1980

CATTAGCGAG ATCGCAGACA TGGTCGCTAC GGCGGTGGGC TACATCGGCG AAACACGTTG 2040

GGATCCAACT AAACCCGATG GAACCCCGCG CAAACTATTG GACGTCTCCG CGCTACGCGA 2100

GTTGGGTTGG CGCCCGCGAA TCGCACTGAA AGACGGCATC GATGCAACGG TGTCGTGGTA 2160

CCGCACAAAT GCCGATGCCG TGAGGAGGTA AAGCTGCGGG CCGGCCGATG TTATCCCTCC 2220

GGCCGGACGG GTAGGGCGAC CTGCCATCGA GTGGTACGGC AGTCGCCTGG CCGGCGAGGC 2280

GCATGGCCTA TGGGAGTATC CCATAGCCTG GCTTGGCTCG CCCCTACGCA TTATCAGTTG 2340

ACCGCTTTCG CGCCAGCTCG CAGGCTCGCG GCAGCATCCC GTTCAGGTCT CCTCATGGTC 2400

CGGTGTGGCA CGACCACGCA AGCTCGAACC GACTCGTTTC CCAATTTCGC ATGCTAATAT 2460

CGCTCGATGG ATTTTTTGCG CAACGCCGGC TTGATGGCTC GTAACGTTAG CACCGAGATG 2520

CTGCGCCACT TCGAACGAAA GCGCCTATTA GTAAACCAAT TCAAAGCATA CGGAGTCAAC 2580

GTTGTTATTG ATGTCGGTGC TAACTCCGGC CAGTTCGGTA GCGCTTTGCG TCGTGCAGGA 2640

TTCAAGAGCC GTATCGTTTC CTTTGAACCT CTTTCGGGGC CATTTGCGCA ACTAACGCGC 2700

GAGTCGGCAT CGGATCCACT ATGGGAGTGT CACCAGTATG CCCTAGGCGA CGCCGATGAG 2760

ACGATTACCA TCAATGTGGC AGGCAATGCG GGGGCAAGTA GTTCCGTGCT GCCGATGCTT 2820

AAAAGTCATC AAGATGCCTT TCCTCCCGCG AATTATATTG GCACCGAAGA CGTTGCAATA 2880

CACCGCCTTG ATTCGGTTGC ATCAGAATTT CTGAACCCTA CCGATGTTAC TTTCCTGAAG 2940

ATCGACGTAC AGGGTTTCGA GAAGCAGGTT ATCGCGGGCA GTAAGTCAAC GCTTAACGAA 3000

AGCTGCGTCG GCATGCAACT CGAACTTTCT TTTATTCCGT TGTACGAAGG TGACATGCTG 3060

ATTCATGAAG CGCTTGAACT TGTCTATTCC CTAGGTTTCA GACTGACGGG TTTGTTGCCC 3120

GGATTTACGG ATCCGCGCAA TGGTCGAATG CTTCAAGCTG ACGGCATTTT CTTCCGTGGG 3180

GACGATTGAC ATAAATGCTT GCGTCGGCAC CCTGCCGGTA TCCAAACGGG CGATCTGGTG 3240

AGCCGGCCTC CCGGGCACCT AATCGACTAT CTAAATTGAG GCGGCCGCGA CGTGCGGCAC 3300

GAACAGGTGG CCGGCTGCTA GCGTTACACA CGTCATGACT GCGCCAGTGT TCTCGATAAT 3360

TATCCCTACC TTCAATGCAG CGGTGACGCT GCAAGCCTGC CTCGGAAGCA TCGTCGGGCA 3420

GACCTACCGG GAAGTGGAAG TGGTCCTTGT CGACGGCGGT TCGACCGATC GGACCCTCGA 3480

CATCGCGAAC AGTTTCCGCC CGGAACTCGG CTCGCGACTG GTCGTTCACA GCGGGCCCGA 3540

TGATGGCCCC TACGACGCCA TGAACCGCGG CGTCGGCGTA GCCACAGGCG AATGGGTACT 3600

TTTTTTAGGC GCCGACGACA CCCTCTACGA ACCAACCACG TTGGCCCAGG TAGCCGCTTT 3660

TCTCGGCGAC CATGCGGCAA GCCATCTTGT CTATGGCGAT GTTGTGATGC GTTCGACGAA 3720

AAGCCGGCAT GCCGGACCTT TCGACCTCGA CCGCCTCCTA TTTGAGACGA ATTTGTGCCA 3780

CCAATCGATC TTTTACCGCC GTGAGCTTTT CGACGGCATC GGCCCTTACA ACCTGCGCTA 3840

CCGAGTCTGG GCGGACTGGG ACTTCAATAT TCGCTGCTTC TCCAACCCGG CGCTGATTAC 3900

CCGCTACATG GACGTCGTGA TTTCCGAATA CAACGACATG ACCGGCTTCA GCATGAGGCA 3960

GGGGACTGAT AAAGAGTTCA GAAAACGGCT GCCAATGTAC TTCTGGGTTG CAGGGTGGGA 4020

GACTTGCAGG CGCATGCTGG CGTTTTTGAA AGACAAGGAG AATCGCCGTC TGGCCTTGCG 4080

TACGCGGTTG ATAAGGGTTA AGGCCGTCTC CAAAGAACGA AGCGCAGAAC CGTAGTCGCG 4140

(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

TTCTACTGCC TGACCTGAGC AGCGCCGAGG CGCGCAGCGC GATCACTGCG ACCTGAATGG 60

CCAGGTGGAA AGCGCCACCG ATCCCGGCAC CGAGTGCCTG ACGATTCGGA TCCCTTGCAC 120

CACAACGAGA GTGAGACCGC CATGATGACG AAATATCGGC TGGGCGGAGT CAACGCCGGA 180

GTGACAAAAG TGAGAACCCG GTGAAGCGAG CGCTTATAAC AGGGATCACG GGGCAGGATG 240

GTTCCTACCT CGCCGAGCTA CTACTGAGCA AGGGATACGA GGTTCACGGG CTCGTTCGTC 300

GAGCTTCGAC GTTTAACACG TCGCGGATCG ATCACCTCTA CGTTGACCCA CACCAACCGG 360

GCGCGCGCTT GTTCTTGCAC TATGCAGACC TCACTGACGG CACCCGGTTG GTGACCCTGC 420

TCAGCAGTAT CGACCCGGAT GAGGTCTACA ACCTCGCAGC GCAGTCCCAT GTGCGCGTCA 480

GCTTTGACGA GCCAGTGCAT ACCGGAGACA CCACCGGCAT GGGATCGATC CGACTTCTGG 540

AAGCAGTCCG CCTTTCTCGG GTGGACTGCC GGTTCTATCA GGCTTCCTCG TCGGAGATGT 600

TCGGCGCATC TCCGCCACCG CAGAACGAAT CGACGCCGTT CTATCCCCGT TCGCCATACG 660

GCGCGGCCAA GGTCTTCTCG TACTGGACGA CTCGCAACTA TCGAGAGGCG TACGGATTAT 720

TCGCAGTGAA TGGCATCTTG TTCAACCATG AGTCCCCCCG GCGCGGCGAG ACTTTCGTGA 780

CCCGAAAGAT CACGCGTGCC GTGGCGCGCA TCCGAGCTGG CTGCCAATCG GAGGTCTATA 840

TGGGCAACCT CGATGCGATC CGCGACTGGG GCTACGCGCC CGAATATGTC GAGGGGATGT 900

GGAGGATGTT GCAAGCGCCT GAACCTGATG ACTACGTCCT GGCGACAGGG CGTGGTTACA 960

CCGTACGTGA GTTCGCTCAA GCTGCTTTTG ACCACGTCGG GCTCGACTGG CAAAAGCACG 1020

TCAAGTTTGA CGACCGCTAT TTGCGCCCCA CCGAGGTCGA TTCGCTAGTA GGAGATGCCG 1080

ACAGGGCGGC CCAGTCACTC GGCTGGAAAG CTTCGGTTCA TACTGGTGAA CTCGCGCGCA 1140

TCATGGTGGA CGCGGACATC GCCGCGTCGG AGTGCGATGG CACACCATGG ATCGACACGC 1200

CGATGTTGCC TGGTTGGGGC GGAGTAAGTT GACGACTACA CCTGGGCCTC TGGACCGCGC 1260

GATCCACATT GGACTTCTTT AACGCGTTTG CGTCCTGATC CACCTTTCAA CCCCGTTCCG 4200

CGTGACGCGG CGCGCAGAGA GTGGTCGCAT ATCGCGTCAC TGTTCTCGTG CCAGTGCTTG 4260

GAAAGCGTCG AGCACTCTGG TTCGCGTTCT TGACGTTCGC GCCCGCCCCT AGAGGTAGCG 4320

TGTCACGTGA CTGAAGCCAA TGAGTGCAAC TCGGCGTCGC GAAAGGTTTC AGTCGCGGTT 4380

GAGCAAGACA CCGCAAGACT ACTGGAGTGC GTGCACAAGC GCCTCCAGCT CACGG 4435

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 378 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..375

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```
ATG ATC GCT GTG ATC TGG TCG GCG GTG CCG ACA GGA ACC GTC GAC TTG      48
Met Ile Ala Val Ile Trp Ser Ala Val Pro Thr Gly Thr Val Asp Leu
 1               5                  10                  15

TCG ACG ATC ACC TTG TAC CGG TCG ATG TAT GAC CCA ATG TCG TCC GCA      96
Ser Thr Ile Thr Leu Tyr Arg Ser Met Tyr Asp Pro Met Ser Ser Ala
            20                  25                  30

ACC GAG AAG ACG TAC GTC AGG TCC GCC GCC CCG CTT TCA CCC ATG GGC     144
Thr Glu Lys Thr Tyr Val Arg Ser Ala Ala Pro Leu Ser Pro Met Gly
        35                  40                  45

GTC GGG ACG GCG ATG AAA ATG ACG TCC GCG TGC TCG ATT CCG CGT TGC     192
Val Gly Thr Ala Met Lys Met Thr Ser Ala Cys Ser Ile Pro Arg Cys
    50                  55                  60

CGG TCG GTG GTG AAG TCA ATC AGC CCG TTC TCA CGG TTC CTC GCA ATC     240
Arg Ser Val Val Lys Ser Ile Ser Pro Phe Ser Arg Phe Leu Ala Ile
65                  70                  75                  80

AAC TCC CAA CCC GGG CTC GAA AAT CGG GAC ACT GCC TGC GAG GAG CAA     288
Asn Ser Gln Pro Gly Leu Glu Asn Arg Asp Thr Ala Cys Glu Glu Gln
```

```
                  85                  90                  95

ATC GAT CTT GGC CTG ATC GAT ATC GAC ACA GAC GAC ATC GTT GCC GCT       336
Ile Asp Leu Gly Leu Ile Asp Ile Asp Thr Asp Asp Ile Val Ala Ala
            100                 105                 110

ATC CGC GAG ACA GGC GCC CGT GAC GAG GCC TAC ATA GCC TGA             378
Ile Arg Glu Thr Gly Ala Arg Asp Glu Ala Tyr Ile Ala
        115                 120                 125
```

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 125 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

```
Met Ile Ala Val Ile Trp Ser Ala Val Pro Thr Gly Thr Val Asp Leu
  1               5                  10                  15

Ser Thr Ile Thr Leu Tyr Arg Ser Met Tyr Asp Pro Met Ser Ser Ala
             20                  25                  30

Thr Glu Lys Thr Tyr Val Arg Ser Ala Ala Pro Leu Ser Pro Met Gly
         35                  40                  45

Val Gly Thr Ala Met Lys Met Thr Ser Ala Cys Ser Ile Pro Arg Cys
     50                  55                  60

Arg Ser Val Val Lys Ser Ile Ser Pro Phe Ser Arg Phe Leu Ala Ile
 65                  70                  75                  80

Asn Ser Gln Pro Gly Leu Glu Asn Arg Asp Thr Ala Cys Glu Glu Gln
                 85                  90                  95

Ile Asp Leu Gly Leu Ile Asp Ile Asp Thr Asp Asp Ile Val Ala Ala
            100                 105                 110

Ile Arg Glu Thr Gly Ala Arg Asp Glu Ala Tyr Ile Ala
        115                 120                 125
```

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 834 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..831

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

```
GTG TCA TCT GCT CCA ACC GTG TCG GTG ATA ACG ATT TCG CTG AAC GAT        48
Val Ser Ser Ala Pro Thr Val Ser Val Ile Thr Ile Ser Leu Asn Asp
                130                 135                 140

CTC GAG GGA TTG AAA AGC ACC GTG GAG AGC GTT CGC GCG CAG CGC TAT        96
Leu Glu Gly Leu Lys Ser Thr Val Glu Ser Val Arg Ala Gln Arg Tyr
            145                 150                 155

GGG GGG CGA ATC GAG CAC ATC GTC ATC GAC GGT GGA TCG GGC GAC GCC       144
Gly Gly Arg Ile Glu His Ile Val Ile Asp Gly Gly Ser Gly Asp Ala
        160                 165                 170

GTC GTG GAG TAT CTG TCC GGC GAT CCT GGC TTT GCA TAT TGG CAA TCT       192
Val Val Glu Tyr Leu Ser Gly Asp Pro Gly Phe Ala Tyr Trp Gln Ser
    175                 180                 185

CAG CCC GAC AAC GGG AGA TAT GAC GCG ATG AAT CAG GGC ATT GCC CAT       240
Gln Pro Asp Asn Gly Arg Tyr Asp Ala Met Asn Gln Gly Ile Ala His
190                 195                 200                 205

TCG TCG GGC GAC CTG TTG TGG TTT ATG CAC TCC ACG GAT CGT TTC TCC       288
Ser Ser Gly Asp Leu Leu Trp Phe Met His Ser Thr Asp Arg Phe Ser
                210                 215                 220

GAT CCA GAT GCA GTC GCT TCC GTG GTG GAG GCG CTC TCG GGG CAT GGA       336
Asp Pro Asp Ala Val Ala Ser Val Val Glu Ala Leu Ser Gly His Gly
            225                 230                 235

CCA GTA CGT GAT TTG TGG GGT TAC GGG AAA AAC AAC CTT GTC GGA CTC       384
Pro Val Arg Asp Leu Trp Gly Tyr Gly Lys Asn Asn Leu Val Gly Leu
        240                 245                 250

GAC GGC AAA CCA CTT TTC CCT CGG CCG TAC GGC TAT ATG CCG TTT AAG       432
Asp Gly Lys Pro Leu Phe Pro Arg Pro Tyr Gly Tyr Met Pro Phe Lys
    255                 260                 265
```

```
ATG CGG AAA TTT CTG CTC GGC GCG ACG GTT GCG CAT CAG GCG ACA TTC      480
Met Arg Lys Phe Leu Leu Gly Ala Thr Val Ala His Gln Ala Thr Phe
270                 275                 280                 285

TTC GGC GCG TCG CTG GTA GCC AAG TTG GGC GGT TAC GAT CTT GAT TTT      528
Phe Gly Ala Ser Leu Val Ala Lys Leu Gly Gly Tyr Asp Leu Asp Phe
                290                 295                 300

GGA CTC GAG GCG GAC CAG CTG TTC ATC TAC CGT GCC GCA CTA ATA CGG      576
Gly Leu Glu Ala Asp Gln Leu Phe Ile Tyr Arg Ala Ala Leu Ile Arg
            305                 310                 315

CCT CCC GTC ACG ATC GAC CGC GTG GTT TGC GAC TTC GAT GTC ACG GGA      624
Pro Pro Val Thr Ile Asp Arg Val Val Cys Asp Phe Asp Val Thr Gly
        320                 325                 330

CCT GGT TCA ACC CAG CCC ATC CGT GAG CAC TAT CGG ACC CTG CGG CGG      672
Pro Gly Ser Thr Gln Pro Ile Arg Glu His Tyr Arg Thr Leu Arg Arg
    335                 340                 345

CTC TGG GAC CTG CAT GGC GAC TAC CCG CTG GGT GGG CGC AGA GTG TCG      720
Leu Trp Asp Leu His Gly Asp Tyr Pro Leu Gly Gly Arg Arg Val Ser
350                 355                 360                 365

TGG GCT TAC TTG CGT GTG AAG GAG TAC TTG ATT CGG GCC GAC CTG GCC      768
Trp Ala Tyr Leu Arg Val Lys Glu Tyr Leu Ile Arg Ala Asp Leu Ala
                370                 375                 380

GCA TTC AAC GCG GTA AAG TTC TTG CGA GCG AAG TTC GCC AGA GCT TCG      816
Ala Phe Asn Ala Val Lys Phe Leu Arg Ala Lys Phe Ala Arg Ala Ser
            385                 390                 395

CGG AAG CAA AAT TCA TAG                                              834
Arg Lys Gln Asn Ser
        400
```

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 277 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

```
Val Ser Ser Ala Pro Thr Val Ser Val Ile Thr Ile Ser Leu Asn Asp
 1               5                  10                  15

Leu Glu Gly Leu Lys Ser Thr Val Glu Ser Val Arg Ala Gln Arg Tyr
            20                  25                  30

Gly Gly Arg Ile Glu His Ile Val Ile Asp Gly Gly Ser Gly Asp Ala
        35                  40                  45

Val Val Glu Tyr Leu Ser Gly Asp Pro Gly Phe Ala Tyr Trp Gln Ser
    50                  55                  60

Gln Pro Asp Asn Gly Arg Tyr Asp Ala Met Asn Gln Gly Ile Ala His
65                  70                  75                  80

Ser Ser Gly Asp Leu Leu Trp Phe Met His Ser Thr Asp Arg Phe Ser
                85                  90                  95

Asp Pro Asp Ala Val Ala Ser Val Val Glu Ala Leu Ser Gly His Gly
            100                 105                 110

Pro Val Arg Asp Leu Trp Gly Tyr Gly Lys Asn Asn Leu Val Gly Leu
        115                 120                 125

Asp Gly Lys Pro Leu Phe Pro Arg Pro Tyr Gly Tyr Met Pro Phe Lys
    130                 135                 140

Met Arg Lys Phe Leu Leu Gly Ala Thr Val Ala His Gln Ala Thr Phe
145                 150                 155                 160

Phe Gly Ala Ser Leu Val Ala Lys Leu Gly Gly Tyr Asp Leu Asp Phe
                165                 170                 175

Gly Leu Glu Ala Asp Gln Leu Phe Ile Tyr Arg Ala Ala Leu Ile Arg
            180                 185                 190

Pro Pro Val Thr Ile Asp Arg Val Val Cys Asp Phe Asp Val Thr Gly
        195                 200                 205

Pro Gly Ser Thr Gln Pro Ile Arg Glu His Tyr Arg Thr Leu Arg Arg
    210                 215                 220

Leu Trp Asp Leu His Gly Asp Tyr Pro Leu Gly Gly Arg Arg Val Ser
225                 230                 235                 240

Trp Ala Tyr Leu Arg Val Lys Glu Tyr Leu Ile Arg Ala Asp Leu Ala
                245                 250                 255
```

```
Ala Phe Asn Ala Val Lys Phe Leu Arg Ala Lys Phe Ala Arg Ala Ser
            260                 265                 270

Arg Lys Gln Asn Ser
        275
```

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1032 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..1029

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

```
GTG AAG CGA GCG CTT ATA ACA GGG ATC ACG GGG CAG GAT GGT TCC TAC        48
Val Lys Arg Ala Leu Ile Thr Gly Ile Thr Gly Gln Asp Gly Ser Tyr
        280                 285                 290

CTC GCC GAG CTA CTA CTG AGC AAG GGA TAC GAG GTT CAC GGG CTC GTT        96
Leu Ala Glu Leu Leu Leu Ser Lys Gly Tyr Glu Val His Gly Leu Val
    295                 300                 305

CGT CGA GCT TCG ACG TTT AAC ACG TCG CGG ATC GAT CAC CTC TAC GTT       144
Arg Arg Ala Ser Thr Phe Asn Thr Ser Arg Ile Asp His Leu Tyr Val
310                 315                 320                 325

GAC CCA CAC CAA CCG GGC GCG CGC TTG TTC TTG CAC TAT GCA GAC CTC       192
Asp Pro His Gln Pro Gly Ala Arg Leu Phe Leu His Tyr Ala Asp Leu
                330                 335                 340

ACT GAC GGC ACC CGG TTG GTG ACC CTG CTC AGC AGT ATC GAC CCG GAT       240
Thr Asp Gly Thr Arg Leu Val Thr Leu Leu Ser Ser Ile Asp Pro Asp
            345                 350                 355

GAG GTC TAC AAC CTC GCA GCG CAG TCC CAT GTG CGC GTC AGC TTT GAC       288
Glu Val Tyr Asn Leu Ala Ala Gln Ser His Val Arg Val Ser Phe Asp
        360                 365                 370
```

```
GAG CCA GTG CAT ACC GGA GAC ACC ACC GGC ATG GGA TCG ATC CGA CTT      336
Glu Pro Val His Thr Gly Asp Thr Thr Gly Met Gly Ser Ile Arg Leu
    375                 380                 385

CTG GAA GCA GTC CGC CTT TCT CGG GTG GAC TGC CGG TTC TAT CAG GCT      384
Leu Glu Ala Val Arg Leu Ser Arg Val Asp Cys Arg Phe Tyr Gln Ala
390                 395                 400                 405

TCC TCG TCG GAG ATG TTC GGC GCA TCT CCG CCA CCG CAG AAC GAA TCG      432
Ser Ser Ser Glu Met Phe Gly Ala Ser Pro Pro Pro Gln Asn Glu Ser
                410                 415                 420

ACG CCG TTC TAT CCC CGT TCG CCA TAC GGC GCG GCC AAG GTC TTC TCG      480
Thr Pro Phe Tyr Pro Arg Ser Pro Tyr Gly Ala Ala Lys Val Phe Ser
            425                 430                 435

TAC TGG ACG ACT CGC AAC TAT CGA GAG GCG TAC GGA TTA TTC GCA GTG      528
Tyr Trp Thr Thr Arg Asn Tyr Arg Glu Ala Tyr Gly Leu Phe Ala Val
        440                 445                 450

AAT GGC ATC TTG TTC AAC CAT GAG TCC CCC CGG CGC GGC GAG ACT TTC      576
Asn Gly Ile Leu Phe Asn His Glu Ser Pro Arg Arg Gly Glu Thr Phe
    455                 460                 465

GTG ACC CGA AAG ATC ACG CGT GCC GTG GCG CGC ATC CGA GCT GGC GTC      624
Val Thr Arg Lys Ile Thr Arg Ala Val Ala Arg Ile Arg Ala Gly Val
470                 475                 480                 485

CAA TCG GAG GTC TAT ATG GGC AAC CTC GAT GCG ATC CGC GAC TGG GGC      672
Gln Ser Glu Val Tyr Met Gly Asn Leu Asp Ala Ile Arg Asp Trp Gly
                490                 495                 500

TAC GCG CCC GAA TAT GTC GAG GGG ATG TGG AGG ATG TTG CAA GCG CCT      720
Tyr Ala Pro Glu Tyr Val Glu Gly Met Trp Arg Met Leu Gln Ala Pro
            505                 510                 515

GAA CCT GAT GAC TAC GTC CTG GCG ACA GGG CGT GGT TAC ACC GTA CGT      768
Glu Pro Asp Asp Tyr Val Leu Ala Thr Gly Arg Gly Tyr Thr Val Arg
        520                 525                 530

GAG TTC GCT CAA GCT GCT TTT GAC CAT GTC GGG CTC GAC TGG CAA AAG      816
Glu Phe Ala Gln Ala Ala Phe Asp His Val Gly Leu Asp Trp Gln Lys
    535                 540                 545

CGC GTC AAG TTT GAC GAC CGC TAT TTG CGT CCC ACC GAG GTC GAT TCG      864
Arg Val Lys Phe Asp Asp Arg Tyr Leu Arg Pro Thr Glu Val Asp Ser
550                 555                 560                 565
```

```
CTA GTA GGA GAT GCC GAC AAG GCG GCC CAG TCA CTC GGC TGG AAA GCT      912
Leu Val Gly Asp Ala Asp Lys Ala Ala Gln Ser Leu Gly Trp Lys Ala
                570                 575                 580

TCG GTT CAT ACT GGT GAA CTC GCG CGC ATC ATG GTG GAC GCG GAC ATC      960
Ser Val His Thr Gly Glu Leu Ala Arg Ile Met Val Asp Ala Asp Ile
            585                 590                 595

GCC GCG TTG GAG TGC GAT GGC ACA CCA TGG ATC GAC ACG CCG ATG TTG     1008
Ala Ala Leu Glu Cys Asp Gly Thr Pro Trp Ile Asp Thr Pro Met Leu
        600                 605                 610

CCT GGT TGG GGC AGA GTA AGT TGA                                    1032
Pro Gly Trp Gly Arg Val Ser
    615                 620
```

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 343 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

```
Val Lys Arg Ala Leu Ile Thr Gly Ile Thr Gly Gln Asp Gly Ser Tyr
  1               5                  10                  15

Leu Ala Glu Leu Leu Leu Ser Lys Gly Tyr Glu Val His Gly Leu Val
             20                  25                  30

Arg Arg Ala Ser Thr Phe Asn Thr Ser Arg Ile Asp His Leu Tyr Val
         35                  40                  45

Asp Pro His Gln Pro Gly Ala Arg Leu Phe Leu His Tyr Ala Asp Leu
     50                  55                  60

Thr Asp Gly Thr Arg Leu Val Thr Leu Leu Ser Ser Ile Asp Pro Asp
 65                  70                  75                  80

Glu Val Tyr Asn Leu Ala Ala Gln Ser His Val Arg Val Ser Phe Asp
                 85                  90                  95

Glu Pro Val His Thr Gly Asp Thr Thr Gly Met Gly Ser Ile Arg Leu
            100                 105                 110
```

Leu Glu Ala Val Arg Leu Ser Arg Val Asp Cys Arg Phe Tyr Gln Ala
115 120 125

Ser Ser Ser Glu Met Phe Gly Ala Ser Pro Pro Pro Gln Asn Glu Ser
130 135 140

Thr Pro Phe Tyr Pro Arg Ser Pro Tyr Gly Ala Ala Lys Val Phe Ser
145 150 155 160

Tyr Trp Thr Thr Arg Asn Tyr Arg Glu Ala Tyr Gly Leu Phe Ala Val
165 170 175

Asn Gly Ile Leu Phe Asn His Glu Ser Pro Arg Arg Gly Glu Thr Phe
180 185 190

Val Thr Arg Lys Ile Thr Arg Ala Val Ala Arg Ile Arg Ala Gly Val
195 200 205

Gln Ser Glu Val Tyr Met Gly Asn Leu Asp Ala Ile Arg Asp Trp Gly
210 215 220

Tyr Ala Pro Glu Tyr Val Glu Gly Met Trp Arg Met Leu Gln Ala Pro
225 230 235 240

Glu Pro Asp Asp Tyr Val Leu Ala Thr Gly Arg Gly Tyr Thr Val Arg
245 250 255

Glu Phe Ala Gln Ala Ala Phe Asp His Val Gly Leu Asp Trp Gln Lys
260 265 270

Arg Val Lys Phe Asp Asp Arg Tyr Leu Arg Pro Thr Glu Val Asp Ser
275 280 285

Leu Val Gly Asp Ala Asp Lys Ala Ala Gln Ser Leu Gly Trp Lys Ala
290 295 300

Ser Val His Thr Gly Glu Leu Ala Arg Ile Met Val Asp Ala Asp Ile
305 310 315 320

Ala Ala Leu Glu Cys Asp Gly Thr Pro Trp Ile Asp Thr Pro Met Leu
325 330 335

Pro Gly Trp Gly Arg Val Ser
340

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1032 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..1029

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

```
GTG AAG CGA GCG CTT ATA ACA GGG ATC ACG GGG CAG GAT GGT TCC TAC          48
Val Lys Arg Ala Leu Ile Thr Gly Ile Thr Gly Gln Asp Gly Ser Tyr
345                 350                 355

CTC GCC GAG CTA CTA CTG AGC AAG GGA TAC GAG GTT CAC GGG CTC GTT          96
Leu Ala Glu Leu Leu Leu Ser Lys Gly Tyr Glu Val His Gly Leu Val
360                 365                 370                 375

CGT CGA GCT TCG ACG TTT AAC ACG TCG CGG ATC GAT CAC CTC TAC GTT         144
Arg Arg Ala Ser Thr Phe Asn Thr Ser Arg Ile Asp His Leu Tyr Val
                380                 385                 390

GAC CCA CAC CAA CCG GGC GCG CGC TTG TTC TTG CAC TAT GCA GAC CTC         192
Asp Pro His Gln Pro Gly Ala Arg Leu Phe Leu His Tyr Ala Asp Leu
            395                 400                 405

ACT GAC GGC ACC CGG TTG GTG ACC CTG CTC AGC AGT ATC GAC CCG GAT         240
Thr Asp Gly Thr Arg Leu Val Thr Leu Leu Ser Ser Ile Asp Pro Asp
        410                 415                 420

GAG GTC TAC AAC CTC GCA GCG CAG TCC CAT GTG CGC GTC AGC TTT GAC         288
Glu Val Tyr Asn Leu Ala Ala Gln Ser His Val Arg Val Ser Phe Asp
    425                 430                 435

GAG CCA GTG CAT ACC GGA GAC ACC ACC GGC ATG GGA TCG ATC CGA CTT         336
Glu Pro Val His Thr Gly Asp Thr Thr Gly Met Gly Ser Ile Arg Leu
440                 445                 450                 455

CTG GAA GCA GTC CGC CTT TCT CGG GTG GAC TGC CGG TTC TAT CAG GCT         384
Leu Glu Ala Val Arg Leu Ser Arg Val Asp Cys Arg Phe Tyr Gln Ala
                460                 465                 470

TCC TCG TCG GAG ATG TTC GGC GCA TCT CCG CCA CCG CAG AAC GAA TCG         432
```

```
Ser Ser Ser Glu Met Phe Gly Ala Ser Pro Pro Pro Gln Asn Glu Ser
            475                 480                 485

ACG CCG TTC TAT CCC CGT TCG CCA TAC GGC GCG GCC AAG GTC TTC TCG      480
Thr Pro Phe Tyr Pro Arg Ser Pro Tyr Gly Ala Ala Lys Val Phe Ser
        490                 495                 500

TAC TGG ACG ACT CGC AAC TAT CGA GAG GCG TAC GGA TTA TTC GCA GTG      528
Tyr Trp Thr Thr Arg Asn Tyr Arg Glu Ala Tyr Gly Leu Phe Ala Val
    505                 510                 515

AAT GGC ATC TTG TTC AAC CAT GAG TCC CCC CGG CGC GGC GAG ACT TTC      576
Asn Gly Ile Leu Phe Asn His Glu Ser Pro Arg Arg Gly Glu Thr Phe
520                 525                 530                 535

GTG ACC CGA AAG ATC ACG CGT GCC GTG GCG CGC ATC CGA GCT GGC GTC      624
Val Thr Arg Lys Ile Thr Arg Ala Val Ala Arg Ile Arg Ala Gly Val
                540                 545                 550

CAA TCG GAG GTC TAT ATG GGC AAC CTC GAT GCG ATC CGC GAC TGG GGC      672
Gln Ser Glu Val Tyr Met Gly Asn Leu Asp Ala Ile Arg Asp Trp Gly
            555                 560                 565

TAC GCG CCC GAA TAT GTC GAG GGG ATG TGG AGG ATG TTG CAA GCG CCT      720
Tyr Ala Pro Glu Tyr Val Glu Gly Met Trp Arg Met Leu Gln Ala Pro
        570                 575                 580

GAA CCT GAT GAC TAC GTC CTG GCG ACA GGG CGT GGT TAC ACC GTA CGT      768
Glu Pro Asp Asp Tyr Val Leu Ala Thr Gly Arg Gly Tyr Thr Val Arg
    585                 590                 595

GAG TTC GCT CAA GCT GCT TTT GAC CAC GTC GGG CTC GAC TGG CAA AAG      816
Glu Phe Ala Gln Ala Ala Phe Asp His Val Gly Leu Asp Trp Gln Lys
600                 605                 610                 615

CAC GTC AAG TTT GAC GAC CGC TAT TTG CGC CCC ACC GAG GTC GAT TCG      864
His Val Lys Phe Asp Asp Arg Tyr Leu Arg Pro Thr Glu Val Asp Ser
                620                 625                 630

CTA GTA GGA GAT GCC GAC AGG GCG GCC CAG TCA CTC GGC TGG AAA GCT      912
Leu Val Gly Asp Ala Asp Arg Ala Ala Gln Ser Leu Gly Trp Lys Ala
            635                 640                 645

TCG GTT CAT ACT GGT GAA CTC GCG CGC ATC ATG GTG GAC GCG GAC ATC      960
Ser Val His Thr Gly Glu Leu Ala Arg Ile Met Val Asp Ala Asp Ile
        650                 655                 660

GCC GCG TCG GAG TGC GAT GGC ACA CCA TGG ATC GAC ACG CCG ATG TTG     1008
```

```
Ala Ala Ser Glu Cys Asp Gly Thr Pro Trp Ile Asp Thr Pro Met Leu
    665                 670                 675

CCT GGT TGG GGC GGA GTA AGT TGA                                    1032
Pro Gly Trp Gly Gly Val Ser
680                 685
```

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 343 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

```
Val Lys Arg Ala Leu Ile Thr Gly Ile Thr Gly Gln Asp Gly Ser Tyr
 1               5                  10                  15

Leu Ala Glu Leu Leu Leu Ser Lys Gly Tyr Glu Val His Gly Leu Val
            20                  25                  30

Arg Arg Ala Ser Thr Phe Asn Thr Ser Arg Ile Asp His Leu Tyr Val
        35                  40                  45

Asp Pro His Gln Pro Gly Ala Arg Leu Phe Leu His Tyr Ala Asp Leu
    50                  55                  60

Thr Asp Gly Thr Arg Leu Val Thr Leu Leu Ser Ser Ile Asp Pro Asp
65                  70                  75                  80

Glu Val Tyr Asn Leu Ala Ala Gln Ser His Val Arg Val Ser Phe Asp
                85                  90                  95

Glu Pro Val His Thr Gly Asp Thr Thr Gly Met Gly Ser Ile Arg Leu
            100                 105                 110

Leu Glu Ala Val Arg Leu Ser Arg Val Asp Cys Arg Phe Tyr Gln Ala
        115                 120                 125

Ser Ser Ser Glu Met Phe Gly Ala Ser Pro Pro Pro Gln Asn Glu Ser
    130                 135                 140

Thr Pro Phe Tyr Pro Arg Ser Pro Tyr Gly Ala Ala Lys Val Phe Ser
145                 150                 155                 160
```

```
Tyr Trp Thr Thr Arg Asn Tyr Arg Glu Ala Tyr Gly Leu Phe Ala Val
                165                 170                 175

Asn Gly Ile Leu Phe Asn His Glu Ser Pro Arg Arg Gly Glu Thr Phe
            180                 185                 190

Val Thr Arg Lys Ile Thr Arg Ala Val Ala Arg Ile Arg Ala Gly Val
        195                 200                 205

Gln Ser Glu Val Tyr Met Gly Asn Leu Asp Ala Ile Arg Asp Trp Gly
    210                 215                 220

Tyr Ala Pro Glu Tyr Val Glu Gly Met Trp Arg Met Leu Gln Ala Pro
225                 230                 235                 240

Glu Pro Asp Asp Tyr Val Leu Ala Thr Gly Arg Gly Tyr Thr Val Arg
                245                 250                 255

Glu Phe Ala Gln Ala Ala Phe Asp His Val Gly Leu Asp Trp Gln Lys
            260                 265                 270

His Val Lys Phe Asp Asp Arg Tyr Leu Arg Pro Thr Glu Val Asp Ser
        275                 280                 285

Leu Val Gly Asp Ala Asp Arg Ala Ala Gln Ser Leu Gly Trp Lys Ala
    290                 295                 300

Ser Val His Thr Gly Glu Leu Ala Arg Ile Met Val Asp Ala Asp Ile
305                 310                 315                 320

Ala Ala Ser Glu Cys Asp Gly Thr Pro Trp Ile Asp Thr Pro Met Leu
                325                 330                 335

Pro Gly Trp Gly Gly Val Ser
            340
```

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1020 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

```
GAC GCG TAT GCG ATC GCC AAG ATC GCC GGT ATC CTG CAA GTT CAG GCG        528
Asp Ala Tyr Ala Ile Ala Lys Ile Ala Gly Ile Leu Gln Val Gln Ala
    505                 510                 515

GTT AGG CGC CAA TAT GGG CTG GCG TGG ATC TCT GCG ATG CCG ACT AAC        576
Val Arg Arg Gln Tyr Gly Leu Ala Trp Ile Ser Ala Met Pro Thr Asn
520                 525                 530                 535

CTC TAC GGA CCC GGC GAC AAC TTC TCC CCG TCC GGG TCG CAT CTC TTG        624
Leu Tyr Gly Pro Gly Asp Asn Phe Ser Pro Ser Gly Ser His Leu Leu
                540                 545                 550

CCG GCG CTC ATC CGT CGA TAT GAG GAA GCC AAA GCT GGT GGT GCA GAA        672
Pro Ala Leu Ile Arg Arg Tyr Glu Glu Ala Lys Ala Gly Gly Ala Glu
            555                 560                 565

GAG GTG ACG AAT TGG GGG ACC GGT ACT CCG CGG CGC GAA CTT CTG CAT        720
Glu Val Thr Asn Trp Gly Thr Gly Thr Pro Arg Arg Glu Leu Leu His
        570                 575                 580

GTC GAC GAT CTG GCG AGC GCA TGC CTG TTC CTT TTG GAA CAT TTC GAT        768
Val Asp Asp Leu Ala Ser Ala Cys Leu Phe Leu Leu Glu His Phe Asp
    585                 590                 595

GGT CCG AAC CAC GTC AAC GTG GGC ACC GGC GTC GAT CAC AGC ATT AGC        816
Gly Pro Asn His Val Asn Val Gly Thr Gly Val Asp His Ser Ile Ser
600                 605                 610                 615

GAG ATC GCA GAC ATG GTC GCT ACA GCG GTG GGC TAC ATC GGC GAA ACA        864
Glu Ile Ala Asp Met Val Ala Thr Ala Val Gly Tyr Ile Gly Glu Thr
                620                 625                 630

CGT TGG GAT CCA ACT AAA CCC GAT GGA ACC CCG CGC AAA CTA TTG GAC        912
Arg Trp Asp Pro Thr Lys Pro Asp Gly Thr Pro Arg Lys Leu Leu Asp
            635                 640                 645

GTC TCC GCG CTA CGC GAG TTG GGT TGG CGC CCG CGA ATC GCA CTG AAA        960
Val Ser Ala Leu Arg Glu Leu Gly Trp Arg Pro Arg Ile Ala Leu Lys
        650                 655                 660

GAC GGC ATC GAT GCA ACG GTG TCG TGG TAC CGC ACA AAT GCC GAT GCC       1008
Asp Gly Ile Asp Ala Thr Val Ser Trp Tyr Arg Thr Asn Ala Asp Ala
    665                 670                 675

GTG AGG AGG TAA                                                      1020
Val Arg Arg  *
680
```

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..1020

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

```
GTG CGA TGG CAC ACC ATG GAT CGA CAC GCC GAT GTT GCC TGG TTG GGG          48
Val Arg Trp His Thr Met Asp Arg His Ala Asp Val Ala Trp Leu Gly
    345                 350                 355

CAG AGT AAG TTG ACG ACT ACA CCT GGG CCT CTG GAC CGC GCA ACG CCC          96
Gln Ser Lys Leu Thr Thr Thr Pro Gly Pro Leu Asp Arg Ala Thr Pro
360                 365                 370                 375

GTG TAT ATC GCC GGT CAT CGG GGG CTG GTC GGC TCA GCG CTC GTA CGT         144
Val Tyr Ile Ala Gly His Arg Gly Leu Val Gly Ser Ala Leu Val Arg
                380                 385                 390

AGA TTT GAG GCC GAG GGG TTC ACC AAT CTC ATT GTG CGA TCA CGC GAT         192
Arg Phe Glu Ala Glu Gly Phe Thr Asn Leu Ile Val Arg Ser Arg Asp
            395                 400                 405

GAG ATT GAT CTG ACG GAC CGA GCC GCA ACG TTT GAT TTT GTG TCT GAG         240
Glu Ile Asp Leu Thr Asp Arg Ala Ala Thr Phe Asp Phe Val Ser Glu
        410                 415                 420

ACA AGA CCA CAG GTG ATC ATC GAT GCG GCC GCA CGG GTC GGC GGC ATC         288
Thr Arg Pro Gln Val Ile Ile Asp Ala Ala Ala Arg Val Gly Gly Ile
    425                 430                 435

ATG GCG AAT AAC ACC TAT CCC GCG GAC TTC TTG TCC GAA AAC CTC CGA         336
Met Ala Asn Asn Thr Tyr Pro Ala Asp Phe Leu Ser Glu Asn Leu Arg
440                 445                 450                 455

ATC CAG ACC AAT TTG CTC GAC GCA GCT GTC GCC GTG CGT GTG CCG CGG         384
Ile Gln Thr Asn Leu Leu Asp Ala Ala Val Ala Val Arg Val Pro Arg
                460                 465                 470

CTC CTT TTC CTC GGT TCG TCA TGC ATC TAC CCG AAG TAC GCT CCG CAA         432
Leu Leu Phe Leu Gly Ser Ser Cys Ile Tyr Pro Lys Tyr Ala Pro Gln
            475                 480                 485

CCT ATC CAC GAG AGT GCT TTA TTG ACT GGC CCT TTG GAG CCC ACC AAC         480
Pro Ile His Glu Ser Ala Leu Leu Thr Gly Pro Leu Glu Pro Thr Asn
        490                 495                 500
```

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 340 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

```
Val Arg Trp His Thr Met Asp Arg His Ala Asp Val Ala Trp Leu Gly
 1               5                  10                  15

Gln Ser Lys Leu Thr Thr Thr Pro Gly Pro Leu Asp Arg Ala Thr Pro
            20                  25                  30

Val Tyr Ile Ala Gly His Arg Gly Leu Val Gly Ser Ala Leu Val Arg
        35                  40                  45

Arg Phe Glu Ala Glu Gly Phe Thr Asn Leu Ile Val Arg Ser Arg Asp
    50                  55                  60

Glu Ile Asp Leu Thr Asp Arg Ala Ala Thr Phe Asp Phe Val Ser Glu
65                  70                  75                  80

Thr Arg Pro Gln Val Ile Ile Asp Ala Ala Ala Arg Val Gly Gly Ile
                85                  90                  95

Met Ala Asn Asn Thr Tyr Pro Ala Asp Phe Leu Ser Glu Asn Leu Arg
            100                 105                 110

Ile Gln Thr Asn Leu Leu Asp Ala Ala Val Ala Val Arg Val Pro Arg
        115                 120                 125

Leu Leu Phe Leu Gly Ser Ser Cys Ile Tyr Pro Lys Tyr Ala Pro Gln
    130                 135                 140

Pro Ile His Glu Ser Ala Leu Leu Thr Gly Pro Leu Glu Pro Thr Asn
145                 150                 155                 160

Asp Ala Tyr Ala Ile Ala Lys Ile Ala Gly Ile Leu Gln Val Gln Ala
                165                 170                 175

Val Arg Arg Gln Tyr Gly Leu Ala Trp Ile Ser Ala Met Pro Thr Asn
            180                 185                 190

Leu Tyr Gly Pro Gly Asp Asn Phe Ser Pro Ser Gly Ser His Leu Leu
        195                 200                 205
```

```
Pro Ala Leu Ile Arg Arg Tyr Glu Glu Ala Lys Ala Gly Gly Ala Glu
    210                 215                 220

Glu Val Thr Asn Trp Gly Thr Gly Thr Pro Arg Arg Glu Leu Leu His
225                 230                 235                 240

Val Asp Asp Leu Ala Ser Ala Cys Leu Phe Leu Leu Glu His Phe Asp
                245                 250                 255

Gly Pro Asn His Val Asn Val Gly Thr Gly Val Asp His Ser Ile Ser
            260                 265                 270

Glu Ile Ala Asp Met Val Ala Thr Ala Val Gly Tyr Ile Gly Glu Thr
        275                 280                 285

Arg Trp Asp Pro Thr Lys Pro Asp Gly Thr Pro Arg Lys Leu Leu Asp
    290                 295                 300

Val Ser Ala Leu Arg Glu Leu Gly Trp Arg Pro Arg Ile Ala Leu Lys
305                 310                 315                 320

Asp Gly Ile Asp Ala Thr Val Ser Trp Tyr Arg Thr Asn Ala Asp Ala
                325                 330                 335

Val Arg Arg  *
            340
```

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1020 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..1020

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

```
GTG CGA TGG CAC ACC ATG GAT CGA CAC GCC GAT GTT GCC TGG TTG GGG        48
Val Arg Trp His Thr Met Asp Arg His Ala Asp Val Ala Trp Leu Gly
                345                 350                 355
```

```
CGG AGT AAG TTG ACG ACT ACA CCT GGG CCT CTG GAC CGC GCA ACG CCC          96
Arg Ser Lys Leu Thr Thr Thr Pro Gly Pro Leu Asp Arg Ala Thr Pro
            360                 365                 370

GTG TAT ATC GCC GGT CAT CGG GGG CTG GTC GGC TCA GCG CTC GTA CGT         144
Val Tyr Ile Ala Gly His Arg Gly Leu Val Gly Ser Ala Leu Val Arg
        375                 380                 385

AGA TTT GAG GCC GAG GGG TTC ACC AAT CTC ATT GTG CGA TCA CGC GAT         192
Arg Phe Glu Ala Glu Gly Phe Thr Asn Leu Ile Val Arg Ser Arg Asp
    390                 395                 400

GAG ATT GAT CTG ACG GAC CGA GCC GCA ACG TTT GAT TTT GTG TCT GAG         240
Glu Ile Asp Leu Thr Asp Arg Ala Ala Thr Phe Asp Phe Val Ser Glu
405                 410                 415                 420

ACA AGA CCA CAG GTG ATC ATC GAT GCG GCC GCA CGG GTC GGC GGC ATC         288
Thr Arg Pro Gln Val Ile Ile Asp Ala Ala Ala Arg Val Gly Gly Ile
                425                 430                 435

ATG GCG AAT AAC ACC TAT CCC GCG GAC TTC TTG TCC GAA AAC CTC CGA         336
Met Ala Asn Asn Thr Tyr Pro Ala Asp Phe Leu Ser Glu Asn Leu Arg
            440                 445                 450

ATC CAG ACC AAT TTG CTC GAC GCA GCT GTC GCC GTG CGT GTG CCG CGG         384
Ile Gln Thr Asn Leu Leu Asp Ala Ala Val Ala Val Arg Val Pro Arg
        455                 460                 465

CTC CTT TTC CTC GGT TCG TCA TGC ATC TAC CCG AAG TAC GCT CCG CAA         432
Leu Leu Phe Leu Gly Ser Ser Cys Ile Tyr Pro Lys Tyr Ala Pro Gln
    470                 475                 480

CCT ATC CAC GAG AGT GCT TTA TTG ACT GGC CCT TTG GAG CCC ACC AAC         480
Pro Ile His Glu Ser Ala Leu Leu Thr Gly Pro Leu Glu Pro Thr Asn
485                 490                 495                 500

GAC GCG TAT GCG ATC GCC AAG ATC GCC GGT ATC CTG CAA GTT CAG GCG         528
Asp Ala Tyr Ala Ile Ala Lys Ile Ala Gly Ile Leu Gln Val Gln Ala
                505                 510                 515

GTT AGG CGC CAA TAT GGG CTG GCG TGG ATC TCT GCG ATG CCG ACT AAC         576
Val Arg Arg Gln Tyr Gly Leu Ala Trp Ile Ser Ala Met Pro Thr Asn
            520                 525                 530

CTC TAC GGA CCC GGC GAC AAC TTC TCC CCG TCC GGG TCG CAT CTC TTG         624
Leu Tyr Gly Pro Gly Asp Asn Phe Ser Pro Ser Gly Ser His Leu Leu
        535                 540                 545
```

```
CCG GCG CTC ATC CGT CGA TAT GAG GAA GCC AAA GCT GGT GGT GCA GAA       672
Pro Ala Leu Ile Arg Arg Tyr Glu Glu Ala Lys Ala Gly Gly Ala Glu
    550                 555                 560

GAG GTG ACG AAT TGG GGG ACC GGT ACT CCG CGG CGC GAA CTT CTG CAT       720
Glu Val Thr Asn Trp Gly Thr Gly Thr Pro Arg Arg Glu Leu Leu His
565                 570                 575                 580

GTC GAC GAT CTG GCG AGC GCA TGC CTG TTC CTT TTG GAA CAT TTC GAT       768
Val Asp Asp Leu Ala Ser Ala Cys Leu Phe Leu Leu Glu His Phe Asp
                585                 590                 595

GGT CCG AAC CAC GTC AAC GTG GGC ACC GGC GTC GAT CAC AGC ATT AGC       816
Gly Pro Asn His Val Asn Val Gly Thr Gly Val Asp His Ser Ile Ser
            600                 605                 610

GAG ATC GCA GAC ATG GTC GCT ACG GCG GTG GGC TAC ATC GGC GAA ACA       864
Glu Ile Ala Asp Met Val Ala Thr Ala Val Gly Tyr Ile Gly Glu Thr
        615                 620                 625

CGT TGG GAT CCA ACT AAA CCC GAT GGA ACC CCG CGC AAA CTA TTG GAC       912
Arg Trp Asp Pro Thr Lys Pro Asp Gly Thr Pro Arg Lys Leu Leu Asp
    630                 635                 640

GTC TCC GCG CTA CGC GAG TTG GGT TGG CGC CCG CGA ATC GCA CTG AAA       960
Val Ser Ala Leu Arg Glu Leu Gly Trp Arg Pro Arg Ile Ala Leu Lys
645                 650                 655                 660

GAC GGC ATC GAT GCA ACG GTG TCG TGG TAC CGC ACA AAT GCC GAT GCC      1008
Asp Gly Ile Asp Ala Thr Val Ser Trp Tyr Arg Thr Asn Ala Asp Ala
                665                 670                 675

GTG AGG AGG TAA                                                     1020
Val Arg Arg  *
            680
```

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 340 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

```
Val Arg Trp His Thr Met Asp Arg His Ala Asp Val Ala Trp Leu Gly
 1               5                  10                  15

Arg Ser Lys Leu Thr Thr Thr Pro Gly Pro Leu Asp Arg Ala Thr Pro
            20                  25                  30

Val Tyr Ile Ala Gly His Arg Gly Leu Val Gly Ser Ala Leu Val Arg
        35                  40                  45

Arg Phe Glu Ala Glu Gly Phe Thr Asn Leu Ile Val Arg Ser Arg Asp
    50                  55                  60

Glu Ile Asp Leu Thr Asp Arg Ala Ala Thr Phe Asp Phe Val Ser Glu
65                  70                  75                  80

Thr Arg Pro Gln Val Ile Ile Asp Ala Ala Ala Arg Val Gly Gly Ile
                85                  90                  95

Met Ala Asn Asn Thr Tyr Pro Ala Asp Phe Leu Ser Glu Asn Leu Arg
            100                 105                 110

Ile Gln Thr Asn Leu Leu Asp Ala Ala Val Ala Val Arg Val Pro Arg
        115                 120                 125

Leu Leu Phe Leu Gly Ser Ser Cys Ile Tyr Pro Lys Tyr Ala Pro Gln
    130                 135                 140

Pro Ile His Glu Ser Ala Leu Leu Thr Gly Pro Leu Glu Pro Thr Asn
145                 150                 155                 160

Asp Ala Tyr Ala Ile Ala Lys Ile Ala Gly Ile Leu Gln Val Gln Ala
                165                 170                 175

Val Arg Arg Gln Tyr Gly Leu Ala Trp Ile Ser Ala Met Pro Thr Asn
            180                 185                 190

Leu Tyr Gly Pro Gly Asp Asn Phe Ser Pro Ser Gly Ser His Leu Leu
        195                 200                 205

Pro Ala Leu Ile Arg Arg Tyr Glu Glu Ala Lys Ala Gly Gly Ala Glu
    210                 215                 220

Glu Val Thr Asn Trp Gly Thr Gly Thr Pro Arg Arg Glu Leu Leu His
225                 230                 235                 240

Val Asp Asp Leu Ala Ser Ala Cys Leu Phe Leu Leu Glu His Phe Asp
                245                 250                 255
```

```
Gly Pro Asn His Val Asn Val Gly Thr Gly Val Asp His Ser Ile Ser
            260                 265                 270

Glu Ile Ala Asp Met Val Ala Thr Ala Val Gly Tyr Ile Gly Glu Thr
        275                 280                 285

Arg Trp Asp Pro Thr Lys Pro Asp Gly Thr Pro Arg Lys Leu Leu Asp
    290                 295                 300

Val Ser Ala Leu Arg Glu Leu Gly Trp Arg Pro Arg Ile Ala Leu Lys
305                 310                 315                 320

Asp Gly Ile Asp Ala Thr Val Ser Trp Tyr Arg Thr Asn Ala Asp Ala
                325                 330                 335

Val Arg Arg  *
            340
```

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 723 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..720

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

```
ATG GAT TTT TTG CGC AAC GCC GGC TTG ATG GCT CGT AAC GTT AGT ACC        48
Met Asp Phe Leu Arg Asn Ala Gly Leu Met Ala Arg Asn Val Ser Thr
                345                 350                 355

GAG ATG CTG CGC CAC TTC GAA CGA AAG CGC CTA TTA GTA AAC CAA TTC        96
Glu Met Leu Arg His Phe Glu Arg Lys Arg Leu Leu Val Asn Gln Phe
            360                 365                 370

AAA GCA TAC GGA GTC AAC GTT GTT ATT GAT GTC GGT GCT AAC TCC GGC       144
Lys Ala Tyr Gly Val Asn Val Val Ile Asp Val Gly Ala Asn Ser Gly
        375                 380                 385
```

```
CAG TTC GGT AGC GCT TTG CGT CGT GCA GGA TTC AAG AGC CGT ATC GTT        192
Gln Phe Gly Ser Ala Leu Arg Arg Ala Gly Phe Lys Ser Arg Ile Val
    390                 395                 400

TCC TTT GAA CCT CTT TCG GGG CCA TTT GCG CAA CTA ACG CGC AAG TCG        240
Ser Phe Glu Pro Leu Ser Gly Pro Phe Ala Gln Leu Thr Arg Lys Ser
405                 410                 415                 420

GCA TCG GAT CCA CTA TGG GAG TGT CAC CAG TAT GCC CTA GGC GAC GCC        288
Ala Ser Asp Pro Leu Trp Glu Cys His Gln Tyr Ala Leu Gly Asp Ala
                425                 430                 435

GAT GAG ACG ATT ACC ATC AAT GTG GCA GGC AAT GCG GGG GCA AGT AGT        336
Asp Glu Thr Ile Thr Ile Asn Val Ala Gly Asn Ala Gly Ala Ser Ser
            440                 445                 450

TCC GTG CTG CCG ATG CTT AAA AGT CAT CAA GAT GCC TTT CCT CCC GCG        384
Ser Val Leu Pro Met Leu Lys Ser His Gln Asp Ala Phe Pro Pro Ala
        455                 460                 465

AAT TAT ATT GGC ACC GAA GAC GTT GCA ATA CAC CGC CTT GAT TCG GTT        432
Asn Tyr Ile Gly Thr Glu Asp Val Ala Ile His Arg Leu Asp Ser Val
    470                 475                 480

GCA TCA GAA TTT CTG AAC CCT ACC GAT GTT ACT TTC CTG AAG ATC GAC        480
Ala Ser Glu Phe Leu Asn Pro Thr Asp Val Thr Phe Leu Lys Ile Asp
485                 490                 495                 500

GTA CAG GGT TTC GAG AAG CAG GTT ATC ACG GGC AGT AAG TCA ACG CTT        528
Val Gln Gly Phe Glu Lys Gln Val Ile Thr Gly Ser Lys Ser Thr Leu
                505                 510                 515

AAC GAA AGC TGC GTC GGC ATG CAA CTC GAA CTT TCT TTT ATT CCG TTG        576
Asn Glu Ser Cys Val Gly Met Gln Leu Glu Leu Ser Phe Ile Pro Leu
            520                 525                 530

TAC GAA GGT GAC ATG CTG ATT CAT GAA GCG CTT GAA CTT GTC TAT TCC        624
Tyr Glu Gly Asp Met Leu Ile His Glu Ala Leu Glu Leu Val Tyr Ser
        535                 540                 545

CTA GGT TTC AGA CTG ACG GGT TTG TTG CCC GGC TTT ACG GAT CCG CGC        672
Leu Gly Phe Arg Leu Thr Gly Leu Leu Pro Gly Phe Thr Asp Pro Arg
    550                 555                 560

AAT GGT CGA ATG CTT CAA GCT GAC GGC ATT TTC TTC CGT GGG GAC GAT        720
Asn Gly Arg Met Leu Gln Ala Asp Gly Ile Phe Phe Arg Gly Asp Asp
565                 570                 575                 580
```

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 240 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

```
Met Asp Phe Leu Arg Asn Ala Gly Leu Met Ala Arg Asn Val Ser Thr
 1               5                  10                  15

Glu Met Leu Arg His Phe Glu Arg Lys Arg Leu Leu Val Asn Gln Phe
            20                  25                  30

Lys Ala Tyr Gly Val Asn Val Val Ile Asp Val Gly Ala Asn Ser Gly
        35                  40                  45

Gln Phe Gly Ser Ala Leu Arg Arg Ala Gly Phe Lys Ser Arg Ile Val
    50                  55                  60

Ser Phe Glu Pro Leu Ser Gly Pro Phe Ala Gln Leu Thr Arg Lys Ser
65                  70                  75                  80

Ala Ser Asp Pro Leu Trp Glu Cys His Gln Tyr Ala Leu Gly Asp Ala
                85                  90                  95

Asp Glu Thr Ile Thr Ile Asn Val Ala Gly Asn Ala Gly Ala Ser Ser
            100                 105                 110

Ser Val Leu Pro Met Leu Lys Ser His Gln Asp Ala Phe Pro Pro Ala
        115                 120                 125

Asn Tyr Ile Gly Thr Glu Asp Val Ala Ile His Arg Leu Asp Ser Val
    130                 135                 140

Ala Ser Glu Phe Leu Asn Pro Thr Asp Val Thr Phe Leu Lys Ile Asp
145                 150                 155                 160

Val Gln Gly Phe Glu Lys Gln Val Ile Thr Gly Ser Lys Ser Thr Leu
                165                 170                 175

Asn Glu Ser Cys Val Gly Met Gln Leu Glu Leu Ser Phe Ile Pro Leu
            180                 185                 190
```

```
Tyr Glu Gly Asp Met Leu Ile His Glu Ala Leu Glu Leu Val Tyr Ser
        195                 200                 205

Leu Gly Phe Arg Leu Thr Gly Leu Leu Pro Gly Phe Thr Asp Pro Arg
    210                 215                 220

Asn Gly Arg Met Leu Gln Ala Asp Gly Ile Phe Phe Arg Gly Asp Asp
225                 230                 235                 240
```

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 723 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..720

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

```
ATG GAT TTT TTG CGC AAC GCC GGC TTG ATG GCT CGT AAC GTT AGC ACC        48
Met Asp Phe Leu Arg Asn Ala Gly Leu Met Ala Arg Asn Val Ser Thr
                245                 250                 255

GAG ATG CTG CGC CAC TTC GAA CGA AAG CGC CTA TTA GTA AAC CAA TTC        96
Glu Met Leu Arg His Phe Glu Arg Lys Arg Leu Leu Val Asn Gln Phe
            260                 265                 270

AAA GCA TAC GGA GTC AAC GTT GTT ATT GAT GTC GGT GCT AAC TCC GGC       144
Lys Ala Tyr Gly Val Asn Val Val Ile Asp Val Gly Ala Asn Ser Gly
        275                 280                 285

CAG TTC GGT AGC GCT TTG CGT CGT GCA GGA TTC AAG AGC CGT ATC GTT       192
Gln Phe Gly Ser Ala Leu Arg Arg Ala Gly Phe Lys Ser Arg Ile Val
    290                 295                 300

TCC TTT GAA CCT CTT TCG GGG CCA TTT GCG CAA CTA ACG CGC GAG TCG       240
Ser Phe Glu Pro Leu Ser Gly Pro Phe Ala Gln Leu Thr Arg Glu Ser
305                 310                 315                 320
```

```
GCA TCG GAT CCA CTA TGG GAG TGT CAC CAG TAT GCC CTA GGC GAC GCC      288
Ala Ser Asp Pro Leu Trp Glu Cys His Gln Tyr Ala Leu Gly Asp Ala
                325                 330                 335

GAT GAG ACG ATT ACC ATC AAT GTG GCA GGC AAT GCG GGG GCA AGT AGT      336
Asp Glu Thr Ile Thr Ile Asn Val Ala Gly Asn Ala Gly Ala Ser Ser
            340                 345                 350

TCC GTG CTG CCG ATG CTT AAA AGT CAT CAA GAT GCC TTT CCT CCC GCG      384
Ser Val Leu Pro Met Leu Lys Ser His Gln Asp Ala Phe Pro Pro Ala
        355                 360                 365

AAT TAT ATT GGC ACC GAA GAC GTT GCA ATA CAC CGC CTT GAT TCG GTT      432
Asn Tyr Ile Gly Thr Glu Asp Val Ala Ile His Arg Leu Asp Ser Val
    370                 375                 380

GCA TCA GAA TTT CTG AAC CCT ACC GAT GTT ACT TTC CTG AAG ATC GAC      480
Ala Ser Glu Phe Leu Asn Pro Thr Asp Val Thr Phe Leu Lys Ile Asp
385                 390                 395                 400

GTA CAG GGT TTC GAG AAG CAG GTT ATC GCG GGC AGT AAG TCA ACG CTT      528
Val Gln Gly Phe Glu Lys Gln Val Ile Ala Gly Ser Lys Ser Thr Leu
                405                 410                 415

AAC GAA AGC TGC GTC GGC ATG CAA CTC GAA CTT TCT TTT ATT CCG TTG      576
Asn Glu Ser Cys Val Gly Met Gln Leu Glu Leu Ser Phe Ile Pro Leu
            420                 425                 430

TAC GAA GGT GAC ATG CTG ATT CAT GAA GCG CTT GAA CTT GTC TAT TCC      624
Tyr Glu Gly Asp Met Leu Ile His Glu Ala Leu Glu Leu Val Tyr Ser
        435                 440                 445

CTA GGT TTC AGA CTG ACG GGT TTG TTG CCC GGA TTT ACG GAT CCG CGC      672
Leu Gly Phe Arg Leu Thr Gly Leu Leu Pro Gly Phe Thr Asp Pro Arg
    450                 455                 460

AAT GGT CGA ATG CTT CAA GCT GAC GGC ATT TTC TTC CGT GGG GAC GAT      720
Asn Gly Arg Met Leu Gln Ala Asp Gly Ile Phe Phe Arg Gly Asp Asp
465                 470                 475                 480

TGA                                                                  723
```

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 240 amino acids
(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

```
Met Asp Phe Leu Arg Asn Ala Gly Leu Met Ala Arg Asn Val Ser Thr
 1               5                  10                  15

Glu Met Leu Arg His Phe Glu Arg Lys Arg Leu Leu Val Asn Gln Phe
            20                  25                  30

Lys Ala Tyr Gly Val Asn Val Val Ile Asp Val Gly Ala Asn Ser Gly
        35                  40                  45

Gln Phe Gly Ser Ala Leu Arg Arg Ala Gly Phe Lys Ser Arg Ile Val
    50                  55                  60

Ser Phe Glu Pro Leu Ser Gly Pro Phe Ala Gln Leu Thr Arg Glu Ser
65                  70                  75                  80

Ala Ser Asp Pro Leu Trp Glu Cys His Gln Tyr Ala Leu Gly Asp Ala
                85                  90                  95

Asp Glu Thr Ile Thr Ile Asn Val Ala Gly Asn Ala Gly Ala Ser Ser
            100                 105                 110

Ser Val Leu Pro Met Leu Lys Ser His Gln Asp Ala Phe Pro Pro Ala
        115                 120                 125

Asn Tyr Ile Gly Thr Glu Asp Val Ala Ile His Arg Leu Asp Ser Val
    130                 135                 140

Ala Ser Glu Phe Leu Asn Pro Thr Asp Val Thr Phe Leu Lys Ile Asp
145                 150                 155                 160

Val Gln Gly Phe Glu Lys Gln Val Ile Ala Gly Ser Lys Ser Thr Leu
                165                 170                 175

Asn Glu Ser Cys Val Gly Met Gln Leu Glu Leu Ser Phe Ile Pro Leu
            180                 185                 190

Tyr Glu Gly Asp Met Leu Ile His Glu Ala Leu Glu Leu Val Tyr Ser
        195                 200                 205

Leu Gly Phe Arg Leu Thr Gly Leu Leu Pro Gly Phe Thr Asp Pro Arg
    210                 215                 220
```

Asn Gly Arg Met Leu Gln Ala Asp Gly Ile Phe Phe Arg Gly Asp Asp
225 230 235 240

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 801 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..798

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

```
ATG ACT GCG CCA GTG TTC TCG ATA ATT ATC CCT ACC TTC AAT GCA GCG        48
Met Thr Ala Pro Val Phe Ser Ile Ile Ile Pro Thr Phe Asn Ala Ala
                245                 250                 255

GTG ACG CTG CAA GCC TGC CTC GGA AGC ATC GTC GGG CAG ACC TAC CGG        96
Val Thr Leu Gln Ala Cys Leu Gly Ser Ile Val Gly Gln Thr Tyr Arg
            260                 265                 270

GAA GTG GAA GTG GTC CTT GTC GAC GGC GGT TCG ACC GAT CGG ACC CTC       144
Glu Val Glu Val Val Leu Val Asp Gly Gly Ser Thr Asp Arg Thr Leu
        275                 280                 285

GAC ATC GCG AAC AGT TTC CGC CCG GAA CTC GGC TCG CGA CTG GTC GTT       192
Asp Ile Ala Asn Ser Phe Arg Pro Glu Leu Gly Ser Arg Leu Val Val
    290                 295                 300

CAC AGC GGG CCC GAT GAT GGC CCC TAC GAC GCC ATG AAC CGC GGC GTC       240
His Ser Gly Pro Asp Asp Gly Pro Tyr Asp Ala Met Asn Arg Gly Val
305                 310                 315                 320

GGC GTG GCC ACA GGC GAA TGG GTA CTT TTT TTA GGC GCC GAC GAC ACC       288
Gly Val Ala Thr Gly Glu Trp Val Leu Phe Leu Gly Ala Asp Asp Thr
                325                 330                 335

CTC TAC GAA CCA ACC ACG TTG GCC CAG GTA GCC GCT TTT CTC GGC GAC       336
Leu Tyr Glu Pro Thr Thr Leu Ala Gln Val Ala Ala Phe Leu Gly Asp
```

```
                340                 345                 350

CAT GCG GCA AGC CAT CTT GTC TAT GGC GAT GTT GTG ATG CGT TCG ACG      384
His Ala Ala Ser His Leu Val Tyr Gly Asp Val Val Met Arg Ser Thr
        355                 360                 365

AAA AGC CGG CAT GCC GGA CCT TTC GAC CTC GAC CGC CTC CTA TTT GAG      432
Lys Ser Arg His Ala Gly Pro Phe Asp Leu Asp Arg Leu Leu Phe Glu
    370                 375                 380

ACG AAT TTG TGC CAC CAA TCG ATC TTT TAC CGC CGT GAG CTT TTC GAC      480
Thr Asn Leu Cys His Gln Ser Ile Phe Tyr Arg Arg Glu Leu Phe Asp
385                 390                 395                 400

GGC ATC GGC CCT TAC AAC CTG CGC TAC CGA GTC TGG GCG GAC TGG GAC      528
Gly Ile Gly Pro Tyr Asn Leu Arg Tyr Arg Val Trp Ala Asp Trp Asp
                405                 410                 415

TTC AAT ATT CGC TGC TTC TCC AAC CCG GCG CTG ATT ACC CGC TAC ATG      576
Phe Asn Ile Arg Cys Phe Ser Asn Pro Ala Leu Ile Thr Arg Tyr Met
            420                 425                 430

GAC GTC GTG ATT TCC GAA TAC AAC GAC ATG ACC GGC TTC AGC ATG AGG      624
Asp Val Val Ile Ser Glu Tyr Asn Asp Met Thr Gly Phe Ser Met Arg
        435                 440                 445

CAG GGG ACT GAT AAA GAG TTC AGA AAA CGG CTG CCA ATG TAC TTC TGG      672
Gln Gly Thr Asp Lys Glu Phe Arg Lys Arg Leu Pro Met Tyr Phe Trp
    450                 455                 460

GTT GCA GGG TGG GAG ACT TGC AGG CGC ATG CTG GCG TTT TTG AAA GAC      720
Val Ala Gly Trp Glu Thr Cys Arg Arg Met Leu Ala Phe Leu Lys Asp
465                 470                 475                 480

AAG GAG AAT CGC CGT CTG GCC TTG CGT ACG CGG TTG ATA AGG GTT AAG      768
Lys Glu Asn Arg Arg Leu Ala Leu Arg Thr Arg Leu Ile Arg Val Lys
                485                 490                 495

GCC GTC TCC AAA GAA CGA AGC GCA GAA CCG TAG                          801
Ala Val Ser Lys Glu Arg Ser Ala Glu Pro
            500                 505
```

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 266 amino acids
(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

```
Met Thr Ala Pro Val Phe Ser Ile Ile Ile Pro Thr Phe Asn Ala Ala
 1               5                  10                  15

Val Thr Leu Gln Ala Cys Leu Gly Ser Ile Val Gly Gln Thr Tyr Arg
            20                  25                  30

Glu Val Glu Val Val Leu Val Asp Gly Gly Ser Thr Asp Arg Thr Leu
        35                  40                  45

Asp Ile Ala Asn Ser Phe Arg Pro Glu Leu Gly Ser Arg Leu Val Val
    50                  55                  60

His Ser Gly Pro Asp Asp Gly Pro Tyr Asp Ala Met Asn Arg Gly Val
65                  70                  75                  80

Gly Val Ala Thr Gly Glu Trp Val Leu Phe Leu Gly Ala Asp Asp Thr
                85                  90                  95

Leu Tyr Glu Pro Thr Thr Leu Ala Gln Val Ala Ala Phe Leu Gly Asp
            100                 105                 110

His Ala Ala Ser His Leu Val Tyr Gly Asp Val Val Met Arg Ser Thr
        115                 120                 125

Lys Ser Arg His Ala Gly Pro Phe Asp Leu Asp Arg Leu Leu Phe Glu
    130                 135                 140

Thr Asn Leu Cys His Gln Ser Ile Phe Tyr Arg Arg Glu Leu Phe Asp
145                 150                 155                 160

Gly Ile Gly Pro Tyr Asn Leu Arg Tyr Arg Val Trp Ala Asp Trp Asp
                165                 170                 175

Phe Asn Ile Arg Cys Phe Ser Asn Pro Ala Leu Ile Thr Arg Tyr Met
            180                 185                 190

Asp Val Val Ile Ser Glu Tyr Asn Asp Met Thr Gly Phe Ser Met Arg
        195                 200                 205

Gln Gly Thr Asp Lys Glu Phe Arg Lys Arg Leu Pro Met Tyr Phe Trp
    210                 215                 220

Val Ala Gly Trp Glu Thr Cys Arg Arg Met Leu Ala Phe Leu Lys Asp
```

```
225                 230                 235                 240

Lys Glu Asn Arg Arg Leu Ala Leu Arg Thr Arg Leu Ile Arg Val Lys
                245                 250                 255

Ala Val Ser Lys Glu Arg Ser Ala Glu Pro
            260                 265
```

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 801 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..798

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

```
ATG ACT GCG CCA GTG TTC TCG ATA ATT ATC CCT ACC TTC AAT GCA GCG        48
Met Thr Ala Pro Val Phe Ser Ile Ile Ile Pro Thr Phe Asn Ala Ala
            270                 275                 280

GTG ACG CTG CAA GCC TGC CTC GGA AGC ATC GTC GGG CAG ACC TAC CGG        96
Val Thr Leu Gln Ala Cys Leu Gly Ser Ile Val Gly Gln Thr Tyr Arg
        285                 290                 295

GAA GTG GAA GTG GTC CTT GTC GAC GGC GGT TCG ACC GAT CGG ACC CTC       144
Glu Val Glu Val Val Leu Val Asp Gly Gly Ser Thr Asp Arg Thr Leu
    300                 305                 310

GAC ATC GCG AAC AGT TTC CGC CCG GAA CTC GGC TCG CGA CTG GTC GTT       192
Asp Ile Ala Asn Ser Phe Arg Pro Glu Leu Gly Ser Arg Leu Val Val
315                 320                 325                 330

CAC AGC GGG CCC GAT GAT GGC CCC TAC GAC GCC ATG AAC CGC GGC GTC       240
His Ser Gly Pro Asp Asp Gly Pro Tyr Asp Ala Met Asn Arg Gly Val
                335                 340                 345

GGC GTA GCC ACA GGC GAA TGG GTA CTT TTT TTA GGC GCC GAC GAC ACC       288
Gly Val Ala Thr Gly Glu Trp Val Leu Phe Leu Gly Ala Asp Asp Thr
```

```
                350                 355                 360

CTC TAC GAA CCA ACC ACG TTG GCC CAG GTA GCC GCT TTT CTC GGC GAC      336
Leu Tyr Glu Pro Thr Thr Leu Ala Gln Val Ala Ala Phe Leu Gly Asp
        365                 370                 375

CAT GCG GCA AGC CAT CTT GTC TAT GGC GAT GTT GTG ATG CGT TCG ACG      384
His Ala Ala Ser His Leu Val Tyr Gly Asp Val Val Met Arg Ser Thr
    380                 385                 390

AAA AGC CGG CAT GCC GGA CCT TTC GAC CTC GAC CGC CTC CTA TTT GAG      432
Lys Ser Arg His Ala Gly Pro Phe Asp Leu Asp Arg Leu Leu Phe Glu
395                 400                 405                 410

ACG AAT TTG TGC CAC CAA TCG ATC TTT TAC CGC CGT GAG CTT TTC GAC      480
Thr Asn Leu Cys His Gln Ser Ile Phe Tyr Arg Arg Glu Leu Phe Asp
                415                 420                 425

GGC ATC GGC CCT TAC AAC CTG CGC TAC CGA GTC TGG GCG GAC TGG GAC      528
Gly Ile Gly Pro Tyr Asn Leu Arg Tyr Arg Val Trp Ala Asp Trp Asp
            430                 435                 440

TTC AAT ATT CGC TGC TTC TCC AAC CCG GCG CTG ATT ACC CGC TAC ATG      576
Phe Asn Ile Arg Cys Phe Ser Asn Pro Ala Leu Ile Thr Arg Tyr Met
        445                 450                 455

GAC GTC GTG ATT TCC GAA TAC AAC GAC ATG ACC GGC TTC AGC ATG AGG      624
Asp Val Val Ile Ser Glu Tyr Asn Asp Met Thr Gly Phe Ser Met Arg
    460                 465                 470

CAG GGG ACT GAT AAA GAG TTC AGA AAA CGG CTG CCA ATG TAC TTC TGG      672
Gln Gly Thr Asp Lys Glu Phe Arg Lys Arg Leu Pro Met Tyr Phe Trp
475                 480                 485                 490

GTT GCA GGG TGG GAG ACT TGC AGG CGC ATG CTG GCG TTT TTG AAA GAC      720
Val Ala Gly Trp Glu Thr Cys Arg Arg Met Leu Ala Phe Leu Lys Asp
                495                 500                 505

AAG GAG AAT CGC CGT CTG GCC TTG CGT ACG CGG TTG ATA AGG GTT AAG      768
Lys Glu Asn Arg Arg Leu Ala Leu Arg Thr Arg Leu Ile Arg Val Lys
            510                 515                 520

GCC GTC TCC AAA GAA CGA AGC GCA GAA CCG TAG                          801
Ala Val Ser Lys Glu Arg Ser Ala Glu Pro
        525                 530
```

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 266 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

```
Met Thr Ala Pro Val Phe Ser Ile Ile Ile Pro Thr Phe Asn Ala Ala
 1               5                  10                  15

Val Thr Leu Gln Ala Cys Leu Gly Ser Ile Val Gly Gln Thr Tyr Arg
            20                  25                  30

Glu Val Glu Val Val Leu Val Asp Gly Gly Ser Thr Asp Arg Thr Leu
        35                  40                  45

Asp Ile Ala Asn Ser Phe Arg Pro Glu Leu Gly Ser Arg Leu Val Val
    50                  55                  60

His Ser Gly Pro Asp Asp Gly Pro Tyr Asp Ala Met Asn Arg Gly Val
65                  70                  75                  80

Gly Val Ala Thr Gly Glu Trp Val Leu Phe Leu Gly Ala Asp Asp Thr
                85                  90                  95

Leu Tyr Glu Pro Thr Thr Leu Ala Gln Val Ala Ala Phe Leu Gly Asp
            100                 105                 110

His Ala Ala Ser His Leu Val Tyr Gly Asp Val Val Met Arg Ser Thr
        115                 120                 125

Lys Ser Arg His Ala Gly Pro Phe Asp Leu Asp Arg Leu Leu Phe Glu
    130                 135                 140

Thr Asn Leu Cys His Gln Ser Ile Phe Tyr Arg Arg Glu Leu Phe Asp
145                 150                 155                 160

Gly Ile Gly Pro Tyr Asn Leu Arg Tyr Arg Val Trp Ala Asp Trp Asp
                165                 170                 175

Phe Asn Ile Arg Cys Phe Ser Asn Pro Ala Leu Ile Thr Arg Tyr Met
            180                 185                 190

Asp Val Val Ile Ser Glu Tyr Asn Asp Met Thr Gly Phe Ser Met Arg
        195                 200                 205

Gln Gly Thr Asp Lys Glu Phe Arg Lys Arg Leu Pro Met Tyr Phe Trp
```

```
    210                 215                 220

Val Ala Gly Trp Glu Thr Cys Arg Arg Met Leu Ala Phe Leu Lys Asp
225                 230                 235                 240

Lys Glu Asn Arg Arg Leu Ala Leu Arg Thr Arg Leu Ile Arg Val Lys
                245                 250                 255

Ala Val Ser Lys Glu Arg Ser Ala Glu Pro
            260                 265
```

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 867 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..864

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

```
GTG GCC AGC AGA AGT CCC CAC TCC GCT GCG GGT GGT TGG CTA ATT CTT        48
Val Ala Ser Arg Ser Pro His Ser Ala Ala Gly Gly Trp Leu Ile Leu
            270                 275                 280

GGC GGC TCC CTT CTT GTG GTC GGC GTG GCG CAT CCG GTA GGA CTC GCC        96
Gly Gly Ser Leu Leu Val Val Gly Val Ala His Pro Val Gly Leu Ala
        285                 290                 295

GGA GGT GAC GAC GAT GCT GGC GTG GTG CAG CAG CCG ATC GAG GAT GCT       144
Gly Gly Asp Asp Asp Ala Gly Val Val Gln Gln Pro Ile Glu Asp Ala
    300                 305                 310

GGC GGC GGT GGT GTG CTC GGG CAG GAA TCG CCC CCA TTG TTC GAA GGG       192
Gly Gly Gly Gly Val Leu Gly Gln Glu Ser Pro Pro Leu Phe Glu Gly
315                 320                 325                 330

CCA ATG CGA GGC GAT GGC CAG GGA GCG GCG CTC GTA GCC GGC AGC CAC       240
Pro Met Arg Gly Asp Gly Gln Gly Ala Ala Leu Val Ala Gly Ser His
                335                 340                 345
```

```
GAG CCG GAA CAA CAG TTG AGT CCC GGT GTC GTC GAG CGG GGC GAA GCC        288
Glu Pro Glu Gln Gln Leu Ser Pro Gly Val Val Glu Arg Gly Glu Ala
            350                 355                 360

GAT CTC GTC CAA GAT GAC CAG ATC CGC GCG GAG CAG GGT GTC GAT GAT        336
Asp Leu Val Gln Asp Asp Gln Ile Arg Ala Glu Gln Gly Val Asp Asp
        365                 370                 375

CTT GCC GAC GGT GTT GTC GGC CAG GCC GCG GTA GAG GAC CTC GAT CAG        384
Leu Ala Asp Gly Val Val Gly Gln Ala Ala Val Glu Asp Leu Asp Gln
    380                 385                 390

GTC GGC GGC GGT GAA GTA GCG GAC TTT GAA TCC GGC GTG GAC GGC AGC        432
Val Gly Gly Gly Glu Val Ala Asp Phe Glu Ser Gly Val Asp Gly Ser
395                 400                 405                 410

GTG CCC GCA GCC GAT GAG CAG GTG ACT TTT GCC CGT ACC AGG TGG GCC        480
Val Pro Ala Ala Asp Glu Gln Val Thr Phe Ala Arg Thr Arg Trp Ala
                415                 420                 425

AAT GAC CGC CAG GTT CTG TTG TGC CCG AAT CCA TTC CAG GCT CGA CAG        528
Asn Asp Arg Gln Val Leu Leu Cys Pro Asn Pro Phe Gln Ala Arg Gln
            430                 435                 440

GTA GTC GAA CGT GGC TGC GGT GAT CGA CGA TCC GGT GAC GTC GAA CCC        576
Val Val Glu Arg Gly Cys Gly Asp Arg Arg Ser Gly Asp Val Glu Pro
        445                 450                 455

GTC GAG GGT CTT GGT GAC CGG GAA GGC TGC GGC CTT GAG ACG GTT GGC        624
Val Glu Gly Leu Gly Asp Arg Glu Gly Cys Gly Leu Glu Thr Val Gly
    460                 465                 470

GGT GTT GGA GGC ATC GCG GGC AGC GAT CTC GGC CTC AAC CAA CGT CCG        672
Gly Val Gly Gly Ile Ala Gly Ser Asp Leu Gly Leu Asn Gln Arg Pro
475                 480                 485                 490

CAG GAT CTC CTC CGG TGT CCA GCG TTG CGT CTT GGC GAC TTG CAA CAC        720
Gln Asp Leu Leu Arg Cys Pro Ala Leu Arg Leu Gly Asp Leu Gln His
                495                 500                 505

CTC GGC GGC GTT GCG GCG CAC CGT GGC CAG CTT CAA CCG CCG CAG CGC        768
Leu Gly Gly Val Ala Ala His Arg Gly Gln Leu Gln Pro Pro Gln Arg
            510                 515                 520

CGC GTC AAG GTC AGC AGC CAG CGG TGC CGC CGA GGA CGG TGC CAC CGG        816
Arg Val Lys Val Ser Ser Gln Arg Cys Arg Arg Gly Arg Cys His Arg
        525                 530                 535
```

CTT GGC AGC GGT GGT CAT GAG GCC GTC CCG TCG GTG GTG TTG ATC TTG 864
Leu Gly Ser Gly Gly His Glu Ala Val Pro Ser Val Val Leu Ile Leu
540 545 550

TAG 867

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 288 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

Val Ala Ser Arg Ser Pro His Ser Ala Ala Gly Gly Trp Leu Ile Leu
1 5 10 15

Gly Gly Ser Leu Leu Val Val Gly Val Ala His Pro Val Gly Leu Ala
20 25 30

Gly Gly Asp Asp Asp Ala Gly Val Val Gln Gln Pro Ile Glu Asp Ala
35 40 45

Gly Gly Gly Gly Val Leu Gly Gln Glu Ser Pro Pro Leu Phe Glu Gly
50 55 60

Pro Met Arg Gly Asp Gly Gln Gly Ala Ala Leu Val Ala Gly Ser His
65 70 75 80

Glu Pro Glu Gln Gln Leu Ser Pro Gly Val Val Glu Arg Gly Glu Ala
85 90 95

Asp Leu Val Gln Asp Asp Gln Ile Arg Ala Glu Gln Gly Val Asp Asp
100 105 110

Leu Ala Asp Gly Val Val Gly Gln Ala Ala Val Glu Asp Leu Asp Gln
115 120 125

Val Gly Gly Gly Glu Val Ala Asp Phe Glu Ser Gly Val Asp Gly Ser
130 135 140

Val Pro Ala Ala Asp Glu Gln Val Thr Phe Ala Arg Thr Arg Trp Ala
145 150 155 160

Asn Asp Arg Gln Val Leu Leu Cys Pro Asn Pro Phe Gln Ala Arg Gln

```
                165                 170                 175

Val Val Glu Arg Gly Cys Gly Asp Arg Arg Ser Gly Asp Val Glu Pro
            180                 185                 190

Val Glu Gly Leu Gly Asp Arg Glu Gly Cys Gly Leu Glu Thr Val Gly
        195                 200                 205

Gly Val Gly Gly Ile Ala Gly Ser Asp Leu Gly Leu Asn Gln Arg Pro
    210                 215                 220

Gln Asp Leu Leu Arg Cys Pro Ala Leu Arg Leu Gly Asp Leu Gln His
225                 230                 235                 240

Leu Gly Gly Val Ala Ala His Arg Gly Gln Leu Gln Pro Pro Gln Arg
                245                 250                 255

Arg Val Lys Val Ser Ser Gln Arg Cys Arg Arg Gly Arg Cys His Arg
            260                 265                 270

Leu Gly Ser Gly Gly His Glu Ala Val Pro Ser Val Val Leu Ile Leu
        275                 280                 285
```

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1739 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..945

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:945..1736

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

```
ATGGGCTGCC TCAAAGGTGG TGTCGTCGCC AATGTTGTTG TTCCAACACC GGATTATGTG      60
```

CGATTCGCGT CCCACTATGG CTTCGTTCCG GACTTCTGCC ACGGTGCGGA TCCGCAATCG 120

AAGGGCATCG TGGAGAACCT CTGTGGCTAC GCTCAGGACG ACCTTGCGGT GCCGCTGCTG 180

ACCGAAGCTG CGTTAGCCGG TGAGCAGGTC GACCTACGTG CCCTCAACGC CCAGGCGCAA 240

CTATGGTGCG CCGAGGTCAA TGCCACGGTC CACTCGGAGA TCTGCGCCGT GCCCAACGAT 300

CGCTTGGTTG ACGAGCGCAC CGTCTTGAGG GAGCTGCCCT CGCTGCGGCC GACGATCGGC 360

TCGGGGTCGG TGCGCCGTAA GGTCGACGGC CTCTCGTGCA TCCGTTACGG CTCAGCTCGT 420

TACTCGGTGC CTCAGCGGCT CGTCGGTGCC ACCGTGGCGG TGGTGGTCGA TCATGGCGCC 480

CTGATCCTGT TGGAACCTGC GACCGGTGTG ATCGTGGCCG AGCACGAGCT CGTCAGCCCA 540

GGTGAGGTGT CCATCCTCGA TGAACACTAC GACGGACCCA GACCCGCACC CTCGCGTGGT 600

CCTCGCCCGA AAACCCAAGC AGAGAAACGA TTCTGCGCAT TGGGAACCGA AGCGCAGCAG 660

TTCCTCGTCG GTGCTGCTGC GATCGGCAAC ACCCGACTGA AATCCGAACT CGACATTCTG 720

CTCGGCCTTG GCGCCGCCCA CGGCGAACAG GCTTTGATTG ACGCGCTGCG CCGGGCGGTT 780

GCGTTTCGCC GGTTCCGCGC TGCCGACGTG CGCTCGATCC TGGCCGCCGG CGCCGGCACC 840

CCACAACCCC GCCCCGCCGG CGACGCACTC GTGCTCGATC TGCCCACCGT CGAGACCCGC 900

TCGTTGGAGG CCTACAAGAT CAACACCACC GACGGGACGG CCTCATGACC ACCGCTGCCA 960

AGCCGGTGGC ACCGTCCTCG GCGGCACCGC TGGCTGCTGA CCTTGACGCG GCGCTGCGGC 1020

GGTTGAAGCT GGCCACGGTG CGCCGCAACG CCGCCGAGGT GTTGCAAGTC GCCAAGACGC 1080

AACGCTGGAC ACCGGAGGAG ATCCTGCGGA CGTTGGTTGA GGCCGAGATC GCTGCCCGCG 1140

ATGCCTCCAA CACCGCCAAC CGTCTCAAGG CCGCAGCCTT CCCGGTCACC AAGACCCTCG 1200

ACGGGTTCGA CGTCACCGGA TCGTCGATCA CCGCAGCCAC GTTCGACTAC CTGTCGAGCC 1260

TGGAATGGAT TCGGGCACAA CAGAACCTGG CGGTCATTGG CCCACCTGGT ACGGGCAAAA 1320

GTCACCTGCT CATCGGCTGC GGGCACGCTG CCGTCCACGC CGGATTCAAA GTCCGCTACT 1380

TCACCGCCGC CGACCTGATC GAGGTCCTCT ACCGCGGCCT GGCCGACAAC ACCGTCGGCA 1440

AGATCATCGA CACCCTGCTC CGCGCGGATC TGGTCATCTT GGACGAGATC GGCTTCGCCC 1500

CGCTCGACGA CACCGGGACT CAACTGTTGT TCCGGCTCGT GGCTGCCGGC TACGAGCGCC 1560

GCTCCCTGGC CATCGCCTCG CATTGGCCCT TCGAACAATG GGGGCGATTC CTGCCCGAGC 1620

ACACCACCGC CGCCAGCATC CTCGATCGGC TGCTGCACCA CGCCAGCATC GTCGTCACCT 1680

CCGGCGAGTC CTACCGGATG CGCCACGCCG ACCACAAGAA GGGAGCCGCC AAGAATTAG 1739

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 315 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

```
Met Gly Cys Leu Lys Gly Gly Val Val Ala Asn Val Val Val Pro Thr
 1               5                  10                  15

Pro Asp Tyr Val Arg Phe Ala Ser His Tyr Gly Phe Val Pro Asp Phe
            20                  25                  30

Cys His Gly Ala Asp Pro Gln Ser Lys Gly Ile Val Glu Asn Leu Cys
        35                  40                  45

Gly Tyr Ala Gln Asp Asp Leu Ala Val Pro Leu Leu Thr Glu Ala Ala
    50                  55                  60

Leu Ala Gly Glu Gln Val Asp Leu Arg Ala Leu Asn Ala Gln Ala Gln
65                  70                  75                  80

Leu Trp Cys Ala Glu Val Asn Ala Thr Val His Ser Glu Ile Cys Ala
                85                  90                  95

Val Pro Asn Asp Arg Leu Val Asp Glu Arg Thr Val Leu Arg Glu Leu
            100                 105                 110

Pro Ser Leu Arg Pro Thr Ile Gly Ser Gly Ser Val Arg Arg Lys Val
        115                 120                 125

Asp Gly Leu Ser Cys Ile Arg Tyr Gly Ser Ala Arg Tyr Ser Val Pro
    130                 135                 140

Gln Arg Leu Val Gly Ala Thr Val Ala Val Val Val Asp His Gly Ala
145                 150                 155                 160
```

```
Leu Ile Leu Leu Glu Pro Ala Thr Gly Val Ile Val Ala Glu His Glu
                165                 170                 175

Leu Val Ser Pro Gly Glu Val Ser Ile Leu Asp Glu His Tyr Asp Gly
            180                 185                 190

Pro Arg Pro Ala Pro Ser Arg Gly Pro Arg Pro Lys Thr Gln Ala Glu
        195                 200                 205

Lys Arg Phe Cys Ala Leu Gly Thr Glu Ala Gln Gln Phe Leu Val Gly
    210                 215                 220

Ala Ala Ala Ile Gly Asn Thr Arg Leu Lys Ser Glu Leu Asp Ile Leu
225                 230                 235                 240

Leu Gly Leu Gly Ala Ala His Gly Glu Gln Ala Leu Ile Asp Ala Leu
                245                 250                 255

Arg Arg Ala Val Ala Phe Arg Arg Phe Arg Ala Ala Asp Val Arg Ser
            260                 265                 270

Ile Leu Ala Ala Gly Ala Gly Thr Pro Gln Pro Arg Pro Ala Gly Asp
        275                 280                 285

Ala Leu Val Leu Asp Leu Pro Thr Val Glu Thr Arg Ser Leu Glu Ala
    290                 295                 300

Tyr Lys Ile Asn Thr Thr Asp Gly Thr Ala Ser
305                 310                 315
```

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 264 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

```
Met Thr Thr Ala Ala Lys Pro Val Ala Pro Ser Ser Ala Ala Pro Leu
  1                 5                  10                  15

Ala Ala Asp Leu Asp Ala Ala Leu Arg Arg Leu Lys Leu Ala Thr Val
            20                  25                  30

Arg Arg Asn Ala Ala Glu Val Leu Gln Val Ala Lys Thr Gln Arg Trp
```

```
          35                  40                  45

Thr Pro Glu Glu Ile Leu Arg Thr Leu Val Glu Ala Glu Ile Ala Ala
     50                  55                  60

Arg Asp Ala Ser Asn Thr Ala Asn Arg Leu Lys Ala Ala Ala Phe Pro
 65                  70                  75                  80

Val Thr Lys Thr Leu Asp Gly Phe Asp Val Thr Gly Ser Ser Ile Thr
                 85                  90                  95

Ala Ala Thr Phe Asp Tyr Leu Ser Ser Leu Glu Trp Ile Arg Ala Gln
            100                 105                 110

Gln Asn Leu Ala Val Ile Gly Pro Pro Gly Thr Gly Lys Ser His Leu
        115                 120                 125

Leu Ile Gly Cys Gly His Ala Ala Val His Ala Gly Phe Lys Val Arg
    130                 135                 140

Tyr Phe Thr Ala Ala Asp Leu Ile Glu Val Leu Tyr Arg Gly Leu Ala
145                 150                 155                 160

Asp Asn Thr Val Gly Lys Ile Ile Asp Thr Leu Leu Arg Ala Asp Leu
                165                 170                 175

Val Ile Leu Asp Glu Ile Gly Phe Ala Pro Leu Asp Asp Thr Gly Thr
            180                 185                 190

Gln Leu Leu Phe Arg Leu Val Ala Ala Gly Tyr Glu Arg Arg Ser Leu
        195                 200                 205

Ala Ile Ala Ser His Trp Pro Phe Glu Gln Trp Gly Arg Phe Leu Pro
    210                 215                 220

Glu His Thr Thr Ala Ala Ser Ile Leu Asp Arg Leu Leu His His Ala
225                 230                 235                 240

Ser Ile Val Val Thr Ser Gly Glu Ser Tyr Arg Met Arg His Ala Asp
                245                 250                 255

His Lys Lys Gly Ala Ala Lys Asn
            260
```

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 789 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

GTGACGTCTG CTCCGACCGT CTCGGTGATA ACGATCTCGT TCAACGACCT CGACGGGTTG 60

CAGCGCACGG TGAAAAGTGT GCGGGCGCAA CGCTACCGGG GACGCATCGA GCACATCGTA 120

ATCGACGGTG GCAGCGGCGA CGACGTGGTG GCATACCTGT CCGGGTGTGA ACCAGGCTTC 180

GCGTATTGGC AGTCCGAGCC CGACGGCGGG CGGTACGACG CGATGAACCA GGGCATCGCG 240

CACGCATCGG GTGATCTGTT GTGGTTCTTG CACTCCGCCG ATCGTTTTTC CGGGCCCGAC 300

GTGGTAGCCC AGGCCGTGGA GGCGCTATCC GGCAAGGGAC CGGTGTCCGA ATTGTGGGGC 360

TTCGGGATGG ATCGTCTCGT CGGGCTCGAT CGGGTGCGCG GCCCGATACC TTTCAGCCTG 420

CGCAAATTCC TGGCCGGCAA GCAGGTTGTT CCGCATCAAG CATCGTTCTT CGGATCATCG 480

CTGGTGGCCA AGATCGGTGG CTACGACCTT GATTTCGGGA TCGCCGCCGA CCAGGAATTC 540

ATATTGCGGG CCGCGCTGGT ATGCGAGCCG GTCACGATTC GGTGTGTGCT GTGCGAGTTC 600

GACACCACGG GCGTCGGCTC GCACCGGGAA CCAAGCGCGG TCTTCGGTGA TCTGCGCCGC 660

ATGGGCGACC TTCATCGCCG CTACCCGTTC GGGGGAAGGC GAATATCACA TGCCTACCTA 720

CGCGGCCGGG AGTTCTACGC CTACAACAGT CGATTCTGGG AAAACGTCTT CACGCGAATG 780

TCGAAATAG 789

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 262 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

Met Thr Ser Ala Pro Thr Val Ser Val Ile Thr Ile Ser Phe Asn Asp
1 5 10 15

Leu Asp Gly Leu Gln Arg Thr Val Lys Ser Val Arg Ala Gln Arg Tyr
20 25 30

Arg Gly Arg Ile Glu His Ile Val Ile Asp Gly Gly Ser Gly Asp Asp
35 40 45

Val Val Ala Tyr Leu Ser Gly Cys Glu Pro Gly Phe Ala Tyr Trp Gln
50 55 60

Ser Glu Pro Asp Gly Gly Arg Tyr Asp Ala Met Asn Gln Gly Ile Ala
65 70 75 80

His Ala Ser Gly Asp Leu Leu Trp Phe Leu His Ser Ala Asp Arg Phe
85 90 95

Ser Gly Pro Asp Val Val Ala Gln Ala Val Glu Ala Leu Ser Gly Lys
100 105 110

Gly Pro Val Ser Glu Leu Trp Gly Phe Gly Met Asp Arg Leu Val Gly
115 120 125

Leu Asp Arg Val Arg Gly Pro Ile Pro Phe Ser Leu Arg Lys Phe Leu
130 135 140

Ala Gly Lys Gln Val Val Pro His Gln Ala Ser Phe Phe Gly Ser Ser
145 150 155 160

Leu Val Ala Lys Ile Gly Gly Tyr Asp Leu Asp Phe Gly Ile Ala Ala
165 170 175

Asp Gln Glu Phe Ile Leu Arg Ala Ala Leu Val Cys Glu Pro Val Thr
180 185 190

Ile Arg Cys Val Leu Cys Glu Phe Asp Thr Thr Gly Val Gly Ser His
195 200 205

Arg Glu Pro Ser Ala Val Phe Gly Asp Leu Arg Arg Met Gly Asp Leu
210 215 220

His Arg Arg Tyr Pro Phe Gly Gly Arg Arg Ile Ser His Ala Tyr Leu
225 230 235 240

Arg Gly Arg Glu Phe Tyr Ala Tyr Asn Ser Arg Phe Trp Glu Asn Val
245 250 255

GCGGCGCTGG AGTGCGAAGG CAAGCCGTGG ATCGACAAGC CGATGATCGC CGGCCGGACA 1020

TGA 1023

(2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 340 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

Met Lys Arg Ala Leu Ile Thr Gly Ile Thr Gly Gln Asp Gly Ser Tyr
1 5 10 15

Leu Ala Glu Leu Leu Leu Ala Lys Gly Tyr Glu Val His Gly Leu Ile
20 25 30

Arg Arg Ala Ser Thr Phe Asn Thr Ser Arg Ile Asp His Leu Tyr Val
35 40 45

Asp Pro His Gln Pro Gly Ala Arg Leu Phe Leu His Tyr Gly Asp Leu
50 55 60

Ile Asp Gly Thr Arg Leu Val Thr Leu Leu Ser Thr Ile Glu Pro Asp
65 70 75 80

Glu Val Tyr Asn Leu Ala Ala Gln Ser His Val Arg Val Ser Phe Asp
85 90 95

Glu Pro Val His Thr Gly Asp Thr Thr Gly Met Gly Ser Met Arg Leu
100 105 110

Leu Glu Ala Val Arg Leu Ser Arg Val His Cys Arg Phe Tyr Gln Ala
115 120 125

Ser Ser Ser Glu Met Phe Gly Ala Ser Pro Pro Pro Gln Asn Glu Leu
130 135 140

Thr Pro Phe Tyr Pro Arg Ser Pro Tyr Gly Ala Ala Lys Val Tyr Ser
145 150 155 160

Tyr Trp Ala Thr Arg Asn Tyr Arg Glu Ala Tyr Gly Leu Phe Ala Val
165 170 175

```
Asn Gly Ile Leu Phe Asn His Glu Ser Pro Arg Arg Gly Glu Thr Phe
            180                 185                 190

Val Thr Arg Lys Ile Thr Arg Ala Val Ala Arg Ile Lys Ala Gly Ile
        195                 200                 205

Gln Ser Glu Val Tyr Met Gly Asn Leu Asp Ala Val Arg Asp Trp Gly
    210                 215                 220

Tyr Ala Pro Glu Tyr Val Glu Gly Met Trp Arg Met Leu Gln Thr Asp
225                 230                 235                 240

Glu Pro Asp Asp Phe Val Leu Ala Thr Gly Arg Gly Phe Thr Val Arg
                245                 250                 255

Glu Phe Ala Arg Ala Ala Phe Glu His Ala Gly Leu Asp Trp Gln Gln
            260                 265                 270

Tyr Val Lys Phe Asp Gln Arg Tyr Leu Arg Pro Thr Glu Val Asp Ser
        275                 280                 285

Leu Ile Gly Asp Ala Thr Lys Ala Ala Glu Leu Leu Gly Trp Arg Ala
    290                 295                 300

Ser Val His Thr Asp Glu Leu Ala Arg Ile Met Val Asp Ala Asp Met
305                 310                 315                 320

Ala Ala Leu Glu Cys Glu Gly Lys Pro Trp Ile Asp Lys Pro Met Ile
                325                 330                 335

Ala Gly Arg Thr
            340
```

(2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 732 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..729

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

```
ATG AGG CTG GCC CGT CGC GCT CGG AAC ATC TTG CGT CGC AAC GGC ATC          48
Met Arg Leu Ala Arg Arg Ala Arg Asn Ile Leu Arg Arg Asn Gly Ile
                345                 350                 355

GAG GTG TCG CGC TAC TTT GCC GAA CTG GAC TGG GAA CGC AAT TTC TTG          96
Glu Val Ser Arg Tyr Phe Ala Glu Leu Asp Trp Glu Arg Asn Phe Leu
            360                 365                 370

CGC CAA CTG CAA TCG CAT CGG GTC AGT GCC GTG CTC GAT GTC GGG GCC         144
Arg Gln Leu Gln Ser His Arg Val Ser Ala Val Leu Asp Val Gly Ala
        375                 380                 385

AAT TCG GGG CAG TAC GCC AGG GGT CTG CGC GGC GCG GGC TTC GCG GGC         192
Asn Ser Gly Gln Tyr Ala Arg Gly Leu Arg Gly Ala Gly Phe Ala Gly
    390                 395                 400

CGC ATC GTC TCG TTC GAG CCG CTG CCC GGG CCC TTT GCC GTC TTG CAG         240
Arg Ile Val Ser Phe Glu Pro Leu Pro Gly Pro Phe Ala Val Leu Gln
405                 410                 415                 420

CGC AGC GCC TCC ACG GAC CCG TTG TGG GAA TGC CGG CGC TGT GCG CTG         288
Arg Ser Ala Ser Thr Asp Pro Leu Trp Glu Cys Arg Arg Cys Ala Leu
                425                 430                 435

GGC GAT GTC GAT GGA ACC ATC TCG ATC AAC GTC GCC GGC AAC GAG GGC         336
Gly Asp Val Asp Gly Thr Ile Ser Ile Asn Val Ala Gly Asn Glu Gly
            440                 445                 450

GCC AGC AGT TCC GTC TTG CCG ATG TTG AAA CGA CAT CAG GAC GCC TTT         384
Ala Ser Ser Ser Val Leu Pro Met Leu Lys Arg His Gln Asp Ala Phe
        455                 460                 465

CCA CCA GCC AAC TAC GTG GGC GCC CAA CGG GTG CCG ATA CAT CGA CTC         432
Pro Pro Ala Asn Tyr Val Gly Ala Gln Arg Val Pro Ile His Arg Leu
    470                 475                 480

GAT TCC GTG GCT GCA GAC GTT CTG CGG CCC AAC GAT ATT GCG TTC TTG         480
Asp Ser Val Ala Ala Asp Val Leu Arg Pro Asn Asp Ile Ala Phe Leu
485                 490                 495                 500

AAG ATC GAC GTT CAA GGA TTC GAG AAG CAG GTG ATC GCG GGT GGC GAT         528
Lys Ile Asp Val Gln Gly Phe Glu Lys Gln Val Ile Ala Gly Gly Asp
                505                 510                 515

TCA ACG GTG CAC GAC CGA TGC GTC GGC ATG CAG CTC GAG CTG TCT TTC         576
Ser Thr Val His Asp Arg Cys Val Gly Met Gln Leu Glu Leu Ser Phe
```

```
            520                 525                 530

CAG CCG TTG TAC GAG GGT GGC ATG CTC ATC CGC GAG GCG CTC GAT CTC      624
Gln Pro Leu Tyr Glu Gly Gly Met Leu Ile Arg Glu Ala Leu Asp Leu
        535                 540                 545

GTG GAT TCG TTG GGC TTT ACG CTC TCG GGA TTG CAA CCC GGT TTC ACC      672
Val Asp Ser Leu Gly Phe Thr Leu Ser Gly Leu Gln Pro Gly Phe Thr
    550                 555                 560

GAC CCC CGC AAC GGT CGA ATG CTG CAG GCC GAT GGC ATC TTC TTC CGG      720
Asp Pro Arg Asn Gly Arg Met Leu Gln Ala Asp Gly Ile Phe Phe Arg
565                 570                 575                 580

GGC AGC GAT TGA                                                     732
Gly Ser Asp
```

(2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 243 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

```
Met Arg Leu Ala Arg Arg Ala Arg Asn Ile Leu Arg Arg Asn Gly Ile
  1               5                  10                  15

Glu Val Ser Arg Tyr Phe Ala Glu Leu Asp Trp Glu Arg Asn Phe Leu
             20                  25                  30

Arg Gln Leu Gln Ser His Arg Val Ser Ala Val Leu Asp Val Gly Ala
         35                  40                  45

Asn Ser Gly Gln Tyr Ala Arg Gly Leu Arg Gly Ala Gly Phe Ala Gly
     50                  55                  60

Arg Ile Val Ser Phe Glu Pro Leu Pro Gly Pro Phe Ala Val Leu Gln
 65                  70                  75                  80

Arg Ser Ala Ser Thr Asp Pro Leu Trp Glu Cys Arg Arg Cys Ala Leu
                 85                  90                  95

Gly Asp Val Asp Gly Thr Ile Ser Ile Asn Val Ala Gly Asn Glu Gly
```

100 105 110

Ala Ser Ser Ser Val Leu Pro Met Leu Lys Arg His Gln Asp Ala Phe
115 120 125

Pro Pro Ala Asn Tyr Val Gly Ala Gln Arg Val Pro Ile His Arg Leu
130 135 140

Asp Ser Val Ala Ala Asp Val Leu Arg Pro Asn Asp Ile Ala Phe Leu
145 150 155 160

Lys Ile Asp Val Gln Gly Phe Glu Lys Gln Val Ile Ala Gly Gly Asp
165 170 175

Ser Thr Val His Asp Arg Cys Val Gly Met Gln Leu Glu Leu Ser Phe
180 185 190

Gln Pro Leu Tyr Glu Gly Gly Met Leu Ile Arg Glu Ala Leu Asp Leu
195 200 205

Val Asp Ser Leu Gly Phe Thr Leu Ser Gly Leu Gln Pro Gly Phe Thr
210 215 220

Asp Pro Arg Asn Gly Arg Met Leu Gln Ala Asp Gly Ile Phe Phe Arg
225 230 235 240

Gly Ser Asp

(2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 732 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..729

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

GTG AAA TCG TTG AAA CTC GCT CGT TTC ATC GCG CGT AGC GCC GCC TTC 48

```
Val Lys Ser Leu Lys Leu Ala Arg Phe Ile Ala Arg Ser Ala Ala Phe
    245                 250                 255

GAG GTT TCG CGC CGC TAT TCT GAG CGA GAC CTG AAG CAC CAG TTT GTG        96
Glu Val Ser Arg Arg Tyr Ser Glu Arg Asp Leu Lys His Gln Phe Val
260                 265                 270                 275

AAG CAA CTC AAA TCG CGT CGG GTA GAT GTC GTT TTC GAT GTC GGC GCC       144
Lys Gln Leu Lys Ser Arg Arg Val Asp Val Val Phe Asp Val Gly Ala
                280                 285                 290

AAC TCA GGA CAA TAC GCC GCC GGC CTC CGC CGA GCA GCA TAT AAG GGC       192
Asn Ser Gly Gln Tyr Ala Ala Gly Leu Arg Arg Ala Ala Tyr Lys Gly
            295                 300                 305

CGC ATT GTC TCG TTC GAA CCG CTA TCC GGA CCG TTT ACG ATC TTG GAA       240
Arg Ile Val Ser Phe Glu Pro Leu Ser Gly Pro Phe Thr Ile Leu Glu
        310                 315                 320

AGC AAA GCG TCA ACG GAT CCA CTT TGG GAT TGC CGG CAG CAT GCG TTG       288
Ser Lys Ala Ser Thr Asp Pro Leu Trp Asp Cys Arg Gln His Ala Leu
    325                 330                 335

GGC GAT TCT GAT GGA ACG GTT ACG ATC AAT ATC GCA GGA AAC GCC GGT       336
Gly Asp Ser Asp Gly Thr Val Thr Ile Asn Ile Ala Gly Asn Ala Gly
340                 345                 350                 355

CAG AGC AGT TCC GTC TTG CCC ATG CTG AAA AGT CAT CAG AAC GCT TTT       384
Gln Ser Ser Ser Val Leu Pro Met Leu Lys Ser His Gln Asn Ala Phe
                360                 365                 370

CCC CCG GCA AAC TAT GTC GGT ACC CAA GAG GCG TCC ATA CAT CGA CTT       432
Pro Pro Ala Asn Tyr Val Gly Thr Gln Glu Ala Ser Ile His Arg Leu
            375                 380                 385

GAT TCC GTG GCG CCA GAA TTT CTA GGC ATG AAC GGT GTC GCT TTT CTC       480
Asp Ser Val Ala Pro Glu Phe Leu Gly Met Asn Gly Val Ala Phe Leu
        390                 395                 400

AAG GTC GAC GTT CAA GGC TTT GAA AAG CAG GTG CTC GCC GGG GGC AAA       528
Lys Val Asp Val Gln Gly Phe Glu Lys Gln Val Leu Ala Gly Gly Lys
    405                 410                 415

TCA ACC ATA GAT GAC CAT TGC GTC GGC ATG CAA CTC GAA CTG TCC TTC       576
Ser Thr Ile Asp Asp His Cys Val Gly Met Gln Leu Glu Leu Ser Phe
420                 425                 430                 435

CTG CCG TTG TAC GAA GGT GGC ATG CTC ATT CCT GAA GCC CTC GAT CTC       624
```

```
Leu Pro Leu Tyr Glu Gly Gly Met Leu Ile Pro Glu Ala Leu Asp Leu
                440                 445                 450

GTG TAT TCC TTG GGC TTC ACG TTG ACG GGA TTG CTG CCT TGT TTC ATT      672
Val Tyr Ser Leu Gly Phe Thr Leu Thr Gly Leu Leu Pro Cys Phe Ile
            455                 460                 465

GAT GCA AAT AAT GGT CGA ATG TTG CAG GCC GAC GGC ATC TTT TTC CGC      720
Asp Ala Asn Asn Gly Arg Met Leu Gln Ala Asp Gly Ile Phe Phe Arg
        470                 475                 480

GAG GAC GAT TGA                                                     732
Glu Asp Asp
    485
```

(2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 243 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

```
Val Lys Ser Leu Lys Leu Ala Arg Phe Ile Ala Arg Ser Ala Ala Phe
  1               5                  10                  15

Glu Val Ser Arg Arg Tyr Ser Glu Arg Asp Leu Lys His Gln Phe Val
             20                  25                  30

Lys Gln Leu Lys Ser Arg Arg Val Asp Val Val Phe Asp Val Gly Ala
         35                  40                  45

Asn Ser Gly Gln Tyr Ala Ala Gly Leu Arg Arg Ala Ala Tyr Lys Gly
     50                  55                  60

Arg Ile Val Ser Phe Glu Pro Leu Ser Gly Pro Phe Thr Ile Leu Glu
 65                  70                  75                  80

Ser Lys Ala Ser Thr Asp Pro Leu Trp Asp Cys Arg Gln His Ala Leu
                 85                  90                  95

Gly Asp Ser Asp Gly Thr Val Thr Ile Asn Ile Ala Gly Asn Ala Gly
            100                 105                 110

Gln Ser Ser Ser Val Leu Pro Met Leu Lys Ser His Gln Asn Ala Phe
```

```
        115                 120                 125

Pro Pro Ala Asn Tyr Val Gly Thr Gln Glu Ala Ser Ile His Arg Leu
    130                 135                 140

Asp Ser Val Ala Pro Glu Phe Leu Gly Met Asn Gly Val Ala Phe Leu
145                 150                 155                 160

Lys Val Asp Val Gln Gly Phe Glu Lys Gln Val Leu Ala Gly Gly Lys
                165                 170                 175

Ser Thr Ile Asp Asp His Cys Val Gly Met Gln Leu Glu Leu Ser Phe
            180                 185                 190

Leu Pro Leu Tyr Glu Gly Gly Met Leu Ile Pro Glu Ala Leu Asp Leu
        195                 200                 205

Val Tyr Ser Leu Gly Phe Thr Leu Thr Gly Leu Leu Pro Cys Phe Ile
    210                 215                 220

Asp Ala Asn Asn Gly Arg Met Leu Gln Ala Asp Gly Ile Phe Phe Arg
225                 230                 235                 240

Glu Asp Asp
```

(2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 828 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..825

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

```
ATG GTG CAG ACG AAA CGA TAC GCC GGC TTG ACC GCA GCT AAC ACA AAG        48
Met Val Gln Thr Lys Arg Tyr Ala Gly Leu Thr Ala Ala Asn Thr Lys
    245                 250                 255
```

```
AAA GTC GCC ATG GCC GCA CCA ATG TTT TCG ATC ATC ATC CCC ACC TTG          96
Lys Val Ala Met Ala Ala Pro Met Phe Ser Ile Ile Ile Pro Thr Leu
260                 265                 270                 275

AAC GTG GCT GCG GTA TTG CCT GCC TGC CTC GAC AGC ATC GCC CGT CAG         144
Asn Val Ala Ala Val Leu Pro Ala Cys Leu Asp Ser Ile Ala Arg Gln
                280                 285                 290

ACC TGC GGT GAC TTC GAG CTG GTA CTG GTC GAC GGC GGC TCG ACG GAC         192
Thr Cys Gly Asp Phe Glu Leu Val Leu Val Asp Gly Gly Ser Thr Asp
            295                 300                 305

GAA ACC CTC GAC ATC GCC AAC ATT TTC GCC CCC AAC CTC GGC GAG CGG         240
Glu Thr Leu Asp Ile Ala Asn Ile Phe Ala Pro Asn Leu Gly Glu Arg
        310                 315                 320

TTG ATC ATT CAT CGC GAC ACC GAC CAG GGC GTC TAC GAC GCC ATG AAC         288
Leu Ile Ile His Arg Asp Thr Asp Gln Gly Val Tyr Asp Ala Met Asn
    325                 330                 335

CGC GGC GTG GAC CTG GCC ACC GGA ACG TGG TTG CTC TTT CTG GGC GCG         336
Arg Gly Val Asp Leu Ala Thr Gly Thr Trp Leu Leu Phe Leu Gly Ala
340                 345                 350                 355

GAC GAC AGC CTG TAC GAG GCT GAC ACC CTG GCG CGG GTG GCC GCC TTC         384
Asp Asp Ser Leu Tyr Glu Ala Asp Thr Leu Ala Arg Val Ala Ala Phe
                360                 365                 370

ATT GGC GAA CAC GAG CCC AGC GAT CTG GTA TAT GGC GAC GTG ATC ATG         432
Ile Gly Glu His Glu Pro Ser Asp Leu Val Tyr Gly Asp Val Ile Met
            375                 380                 385

CGC TCA ACC AAT TTC CGC TGG GGT GGC GCC TTC GAC CTC GAC CGT CTG         480
Arg Ser Thr Asn Phe Arg Trp Gly Gly Ala Phe Asp Leu Asp Arg Leu
        390                 395                 400

TTG TTC AAG CGC AAC ATC TGC CAT CAG GCG ATC TTC TAC CGC CGC GGA         528
Leu Phe Lys Arg Asn Ile Cys His Gln Ala Ile Phe Tyr Arg Arg Gly
    405                 410                 415

CTC TTC GGC ACC ATC GGT CCC TAC AAC CTC CGC TAC CGG GTC CTG GCC         576
Leu Phe Gly Thr Ile Gly Pro Tyr Asn Leu Arg Tyr Arg Val Leu Ala
420                 425                 430                 435

GAC TGG GAC TTC AAT ATT CGC TGC TTT TCC AAC CCA GCG CTC GTC ACC         624
Asp Trp Asp Phe Asn Ile Arg Cys Phe Ser Asn Pro Ala Leu Val Thr
                440                 445                 450
```

```
CGC TAC ATG CAC GTG GTC GTT GCA AGC TAC AAC GAA TTC GGC GGG CTC      672
Arg Tyr Met His Val Val Val Ala Ser Tyr Asn Glu Phe Gly Gly Leu
            455                 460                 465

AGC AAT ACG ATC GTC GAC AAG GAG TTT TTG AAG CGG CTG CCG ATG TCC      720
Ser Asn Thr Ile Val Asp Lys Glu Phe Leu Lys Arg Leu Pro Met Ser
        470                 475                 480

ACG AGA CTC GGC ATA AGG CTG GTC ATA GTT CTG GTG CGC AGG TGG CCA      768
Thr Arg Leu Gly Ile Arg Leu Val Ile Val Leu Val Arg Arg Trp Pro
    485                 490                 495

AAG GTG ATC AGC AGG GCC ATG GTA ATG CGC ACC GTC ATT TCT TGG CGG      816
Lys Val Ile Ser Arg Ala Met Val Met Arg Thr Val Ile Ser Trp Arg
500                 505                 510                 515

CGC CGA CGT TAG                                                      828
Arg Arg Arg
```

(2) INFORMATION FOR SEQ ID NO: 39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 275 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

```
Met Val Gln Thr Lys Arg Tyr Ala Gly Leu Thr Ala Ala Asn Thr Lys
 1               5                  10                  15

Lys Val Ala Met Ala Ala Pro Met Phe Ser Ile Ile Ile Pro Thr Leu
            20                  25                  30

Asn Val Ala Ala Val Leu Pro Ala Cys Leu Asp Ser Ile Ala Arg Gln
        35                  40                  45

Thr Cys Gly Asp Phe Glu Leu Val Leu Val Asp Gly Gly Ser Thr Asp
    50                  55                  60

Glu Thr Leu Asp Ile Ala Asn Ile Phe Ala Pro Asn Leu Gly Glu Arg
65                  70                  75                  80

Leu Ile Ile His Arg Asp Thr Asp Gln Gly Val Tyr Asp Ala Met Asn
                85                  90                  95
```

Arg Gly Val Asp Leu Ala Thr Gly Thr Trp Leu Leu Phe Leu Gly Ala
100 105 110

Asp Asp Ser Leu Tyr Glu Ala Asp Thr Leu Ala Arg Val Ala Ala Phe
115 120 125

Ile Gly Glu His Glu Pro Ser Asp Leu Val Tyr Gly Asp Val Ile Met
130 135 140

Arg Ser Thr Asn Phe Arg Trp Gly Gly Ala Phe Asp Leu Asp Arg Leu
145 150 155 160

Leu Phe Lys Arg Asn Ile Cys His Gln Ala Ile Phe Tyr Arg Arg Gly
165 170 175

Leu Phe Gly Thr Ile Gly Pro Tyr Asn Leu Arg Tyr Arg Val Leu Ala
180 185 190

Asp Trp Asp Phe Asn Ile Arg Cys Phe Ser Asn Pro Ala Leu Val Thr
195 200 205

Arg Tyr Met His Val Val Val Ala Ser Tyr Asn Glu Phe Gly Gly Leu
210 215 220

Ser Asn Thr Ile Val Asp Lys Glu Phe Leu Lys Arg Leu Pro Met Ser
225 230 235 240

Thr Arg Leu Gly Ile Arg Leu Val Ile Val Leu Val Arg Arg Trp Pro
245 250 255

Lys Val Ile Ser Arg Ala Met Val Met Arg Thr Val Ile Ser Trp Arg
260 265 270

Arg Arg Arg
275

(2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

GATGCCGTGA GGAGGTAAAG CTGC 24

(2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

GATACGGCTC TTGAATCCTG CACG 24

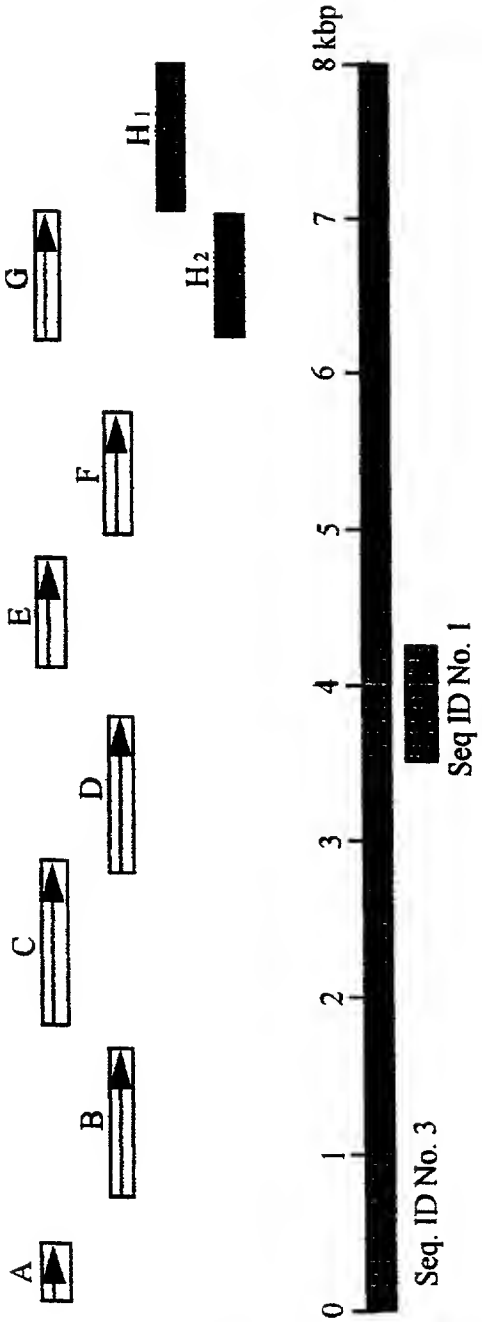

Figure1 a)

Figure1 b)

117-260
N.70283B DMG/IJB/ap

Nixon & Vanderhye P.C. (12/95)

RULE 63 (37 C.F.R. 1.63)
DECLARATION AND POWER OF ATTORNEY
FOR PATENT APPLICATION
IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

As a below named inventor, I hereby declare that my residence, post office address and citizenship are as stated below next to my name, and I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES IN PATHOGENIC MYCOBACTERIA AND THEIR USE AS DIAGNOSTICS, VACCINES AND TARGETS FOR CHEMOTHERAPY

the specification of which (**check** applicable **box(s)**)

- ☐ is attached hereto
- ☐ was filed on **19 June 1998** as U.S. Application Serial No. **(To Be Assigned)** (Atty Dkt. No. 117-260)
- ☒ was filed as PCT International application No. **PCT/GB96/03221** on **23 December 1996**

and (if applicable to U.S. or PCT application) was amended on **22 December 1997**

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above. I acknowledge the duty to disclose information which is material to the patentability of this application in accordance with 37 C.F.R. 1.56. I hereby claim foreign priority benefits under 35 U.S.C. 119/365 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed or, if no priority is claimed, before the filing date of this application:

Priority Foreign Application(s):

Application Number	Country	Day/Month/Year Filed
9526178.0	Great Britain	21 December 1995

I hereby claim the benefit under 35 U.S.C. §119(e) of any United States provisional application(s) listed below.

Application Number	Date/Month/Year Filed

I hereby claim the benefit under 35 U.S.C. 120/365 of all prior United States and PCT International applications listed above or below and, insofar as the subject matter of each of the claims of this application is not disclosed in such prior applications in the manner provided by the first paragraph of 35 U.S.C. 112, I acknowledge the duty to disclose material information as defined in 37 C.F.R 1.56 which occurred between the filing date of the prior applications and the national or PCT international filing date of this application:

Prior U.S./PCT Application(s):

Application Serial No.	Day/Month/Year Filed	Status: patented pending, abandoned
PCT/GB96/03221	23 December 1996	

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon. And I hereby appoint **NIXON & VANDERHYE P.C., 1100 North Glebe Rd., 8th Floor, Arlington, VA 22201-4714, telephone number (703) 816-4000 (to whom all communications are to be directed)**, and the following attorneys thereof (of the same address) individually and collectively my attorneys to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith and with the resulting patent: Arthur R. Crawford, 25327; Larry S. Nixon, 25640; Robert A. Vanderhye, 27076; James T. Hosmer, 30184; Robert W. Faris, 31352; Richard G. Besha, 22770; Mark E. Nusbaum, 32348; Michael J. Keenan, 32106; Bryan H. Davidson, 30251; Stanley C. Spooner, 27393; Leonard C. Mitchard, 29009; Duane M. Byers, 33363; Jeffry H Nelson, 30481; John R. Lastova, 33149; H. Warren Burnam, Jr. 29366; Thomas E. Byrne, 32205; Mary J. Wilson, 32955; J. Scott Davidson, 33489; Alan M. Kagen, 36178, William J. Griffin, 31260; Robert A. Molan, 29834; B. J. Sadoff, 36663; James D. Berquist, 34776; Updeep S Gill, 37334.*

Inventor's Signature: John Herman-Taylor Date: 14th July 1998
Inventor.
John (first) MI HERMON-TAYLOR (last) British (citizenship)
Residence. (city) London (state/country) United Kingdom
Post Office Address: St. George's Hospital Medical School, Dept. Of Surgery, Cranmer Terrace, London, United Kingdom
(Zip Code) SW17 0RE

Inventor's Signature: Tim Doran Date: 5/8/1998
Inventor:
Tim (first) MI DORAN (last) Australian (citizenship)
Residence; (city) Whillington (state/country) Australia
Post Office Address; 1/8 Oxford Street, Whillington, Australia
(Zip Code) VIC 3219

FOR ADDITIONAL INVENTORS, check box ☒ and attach sheet with same information and signature and date for each.

117-260

RULE 63 (37 C.F.R. 1.63)
DECLARATION AND POWER OF ATTORNEY
FOR PATENT APPLICATION
IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Nixon & Vanderhye P.C. (12/95)

3. Inventor's Signature: [signature] Date: 14/8/98
Inventor: Douglas (first) MI MILLAR (last) British (citizenship)
Residence. (city) North Ryde AUV (state/country) Australia
Post Office Address: Csiro Division of Biomolecular Engineering P.O. Box 184, North Ryde, Australia
(Zip Code) NSW 2113

4. Inventor's Signature: [signature] Date: 5/8/98
Inventor: Mark (first) MI TIZARD (last) British (citizenship)
Residence: (city) London GB3 (state/country) United Kingdom
Post Office Address: St. George's Hospital Medical School, Dept. Of Surgery, Cranmer Terrace, London, United Kingdom
(Zip Code) SW17 0RE

5. Inventor's Signature: [signature] Date: 24.7.98
Inventor: Mark (first) MI LOUGHLIN (last) British (citizenship)
Residence. (city) London GB3 (state/country) United Kingdom
Post Office Address: St. George's Hospital Medical School, Dept. Of Surgery, Cranmer Terrace, London, United Kingdom
(Zip Code) SW17 0RE

6. Inventor's Signature: [signature] Date: 14th July 1998
Inventor: Nazira (first) MI SUMAR (last) British (citizenship)
Residence: (city) London GB3 (state/country) United Kingdom
Post Office Address: St. George's Hospital Medical School, Dept. Of Surgery, Cranmer Terrace, London, United Kingdom
(Zip Code) SW17 0RE

7. Inventor's Signature: [signature] Date: 27/7/98
Inventor: John (first) MI FORD (last) British (citizenship)
Residence: (city) London GB3 (state/country) United Kingdom
Post Office Address: St. George's Hospital Medical School, Dept. Of Surgery, Cranmer Terrace, London, United Kingdom
(Zip Code) SW17 0RE

8. Inventor's Signature: _____ Date: _____
Inventor: (first) MI (last) (citizenship)
Residence: (city) (state/country)
Post Office Address:
(Zip Code)

9. Inventor's Signature: _____ Date: _____
Inventor: (first) MI (last) (citizenship)
Residence: (city) (state/country)
Post Office Address:
(Zip Code)

10. Inventor's Signature: _____ Date: _____
Inventor: (first) MI (last) (citizenship)
Residence: (city) (state/country)
Post Office Address.
(Zip Code)

11. Inventor's Signature: _____ Date: _____
Inventor: (first) MI (last) (citizenship)
Residence: (city) (state/country)
Post Office Address:
(Zip Code)